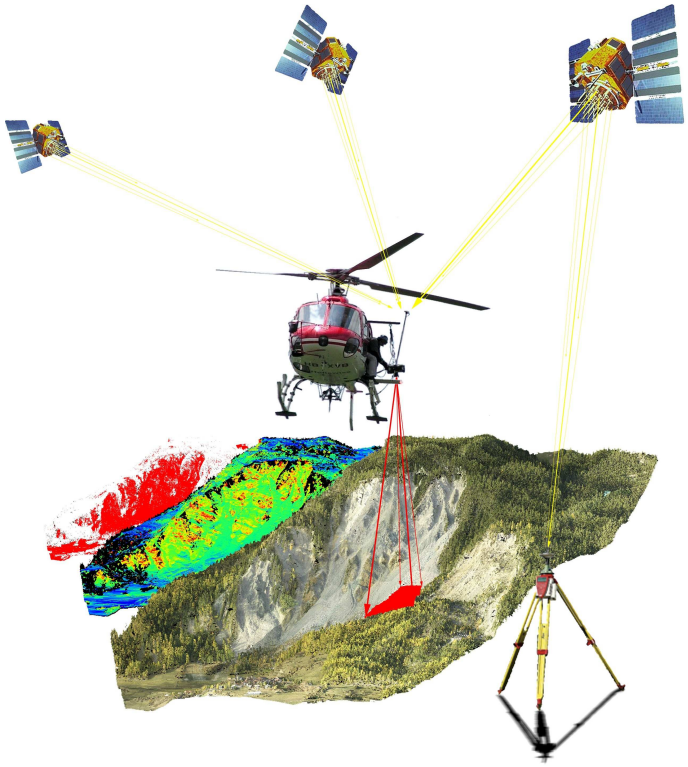
Optical Sensing in Mapping



Jan Skaloud*, EPFL

edition January 2024

* with sections *Frame Cameras* and *Feature Matching* written by Michael Cramer and Norbert Haala, respectively.

Preface

Geographic information systems (GIS) receive data from many sources that are different in technology, geographic coverage, date of capture, and accuracy – to mention few categories. The vast majority of the today's topographical and GIS-data are captured from mobile and possibly autonomous platforms that operate from the air, on the ground (also indoors) or on the water and that are equipped with optical sensors. Although the palette of optical sensors is rather large the most useful for mapping purposes falls into two categories. First are the passive sensors such as digital cameras in frame or line configuration. The main technological concepts of these sensors are introduced in Optical Sensors (Sec. 1) together with Lidar that serves the acquisition of detailed features in cities and terrain structure in natural areas. The optical acquisition is supported by trajectory determination through the combined use of integrated navigation technology, which main concepts are outlined in Navigation Sensors (Sec. 2). The geometrical principals of 3-D restitution of the scene are described first in Photogrammetry (Sec. 3) for the case of frame imagery only, later in Sensor Fusion (Sec. 4) for active sensors and integrated approaches. An overview of Mapping Products (Sec. 5) concludes this manuscript.

Jan Skaloud
Lausanne, January 2023

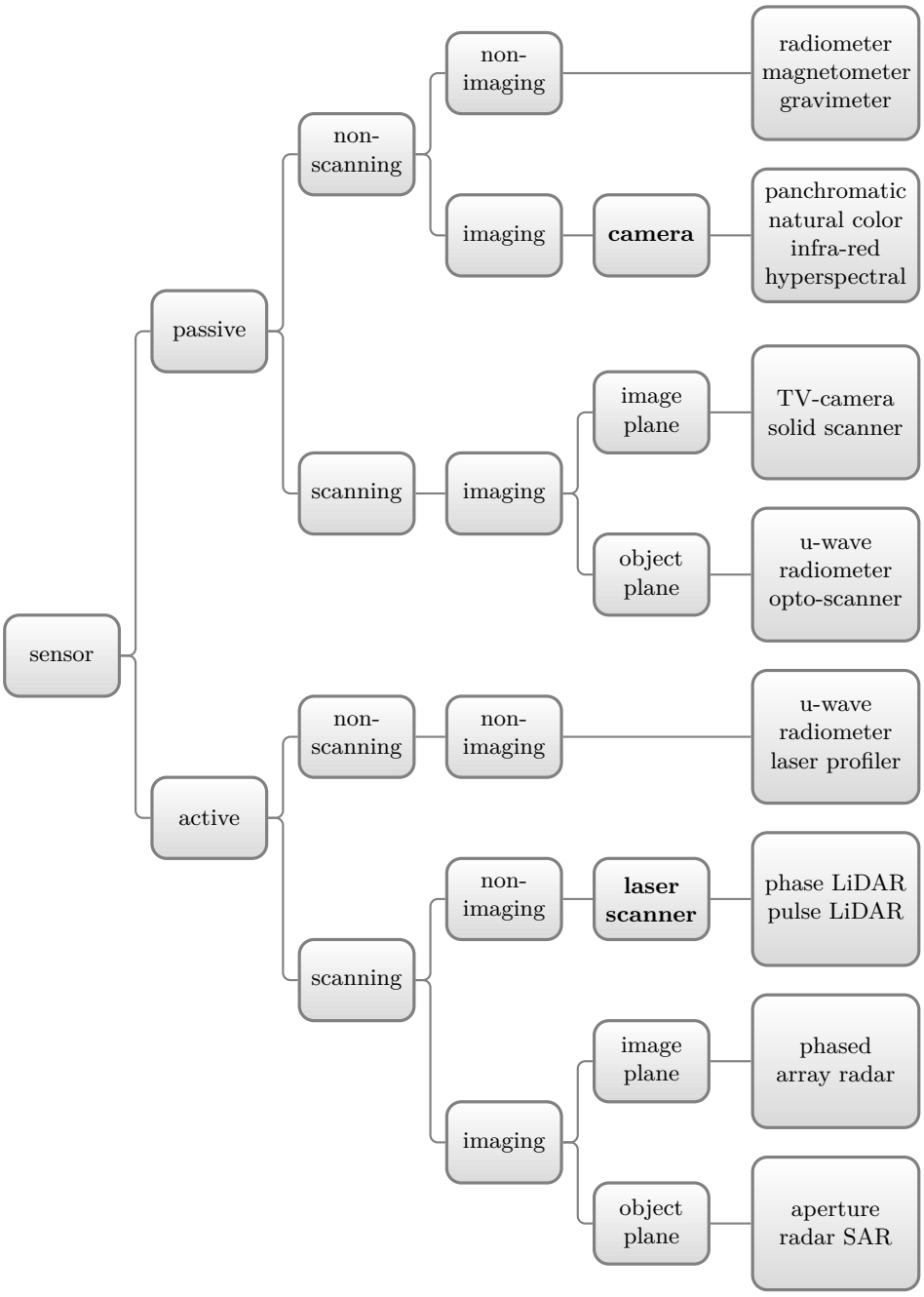


Figure 1: Overview of optical sensors.

Contents

Chapter 1
Optical Sensors

1.1 General remarks on imaging

The acquisition of surface texture is mainly obtained by passive optical sensors that originate from the principles of photography. Photography is a passive method, i.e. the solar energy reflected from the object (or the emitted thermal energy) is recorded by photo sensitive materials or elements. The modern electronic light-sensitive elements are charge-coupled (CCDs) or complementary metal-oxide-semiconductor (CMOS) devices. A CCD array consists of coupled detectors that allow charge to be moved across the array into capacitor bins for further processing. A CMOS detector works independently of neighboring detectors (pixels), as each one has an attached transistor that controls the analog-to-digital conversion and subsequent readout. A CMOS sensor is less expensive to manufacture and has principally faster readout.

Photography in its simplest case is based on the pinhole camera model Wolf (1974). The geometric theory of optical systems assumes straight light rays that have been reflected from an object illuminated by any light source. These rays are entering the camera through the pinhole, forming an inverted image on the plane opposite to the pinhole. This is where the photo sensitive material is placed. In the pinhole camera model, described in detail in Sec. 3.3, each image point is generated by one single light ray passing the pinhole. The resulting 2-D image is an ideal projection of the 3-D objects, since the simple pinhole model neglects for example any distortions and blurring effects due to de-focus. It is assumed that pinhole, object point and image point are defining one straight line (central perspective). The pinhole has the drawback that ideally only one single ray from the bundle of light rays originated from the object point source is forming the corresponding image point. Thus lenses, single or more complex lens systems, are used to enlarge the size of the camera opening but still retaining a focused image. The optical axis of such lens system is defined as the line between incident and emergent nodal point, which are defined in a way that the chief or central rays are passing the lens (system) without deviation, forming the same angles to the optical axis in both nodal points Kraus (2007), i.e. the emerging ray is parallel to the original incident ray (see Fig. 1.1).

The nodal points define the object-space incident and image-space emergent perspective centers. Light rays, entering the lens parallel to the optical

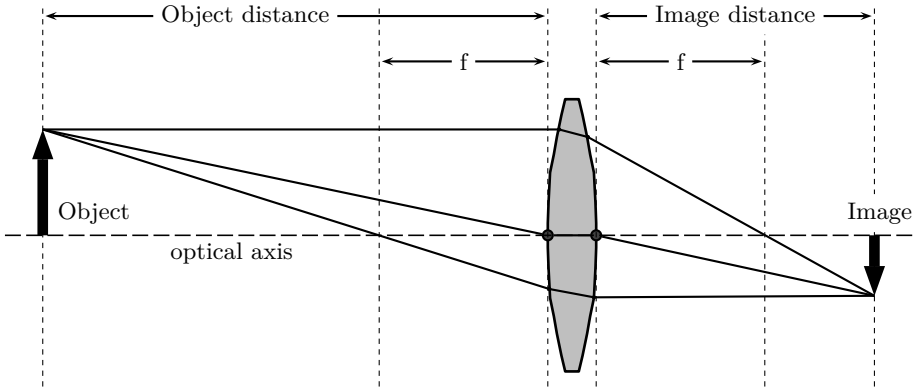


Figure 1.1: Basic imaging principe when using a lens or lens system.

axis come to focus at the focal point. The plane perpendicular to the optical axis, including the focal point, defines the plane of infinite focus or simply the focal plane Wolf (1974). Any parallel rays entering the system come to focus in this plane. The focal length of the system is the distance between the emergent nodal point and the focal plane. The principal point is defined where the optical axis hits the focal plane. The principal point and focal length are the elementary geometrical parameters defining the geometry of a camera. This is called the interior (inner) orientation of the camera. If this interior orientation of the camera and its corresponding image is known, a bundle of image rays can be reconstructed from observed image coordinate measurements. The connection between the bundle of rays towards their correspondences in the image space is expressed by collinearity equation, which is presented in Sec. 3. Reconstruction of 3-D coordinates of objects derived exclusively from 2-D image observations is detailed in Sec. 3, while the benefit of integrating navigation sensors into this process is described in Sec. 4.

1.2 Frame cameras

Until recently the acquisition of texture was exclusively done using films. They are now almost completely replaced by electronic sensors at least for consumer-grade photography. Consequently and similar to consumer grade photography, analogue film based systems are rapidly phasing out in operational photogrammetric environments and digital sensors to a large extent have replaced their analogue predecessors. According to its geometry the airborne digital cameras fall into the two large categories of frame and line cameras, where the latter are also referred to as push-broom sensors. The concept of both is described in sequel.

1.2.1 Single head

Airborne or satellite platforms employs large image formats to guarantee an efficient data acquisition as the available image size directly influences the cost of covering a certain area with imagery. Indeed, a smaller image format requires more images to record a given scene with the same spatial resolution. Especially in airborne imaging this negatively influences the efficiency of image data recording and processing. Therefore, traditional analogue mapping cameras have been designed with large formats of about 23 x 23 cm^2. For those, focal lengths of 30, 15 or 8 cm are utilized depending on the needed field of view (FOV), which is 60, 95 and 125 degrees, respectively.

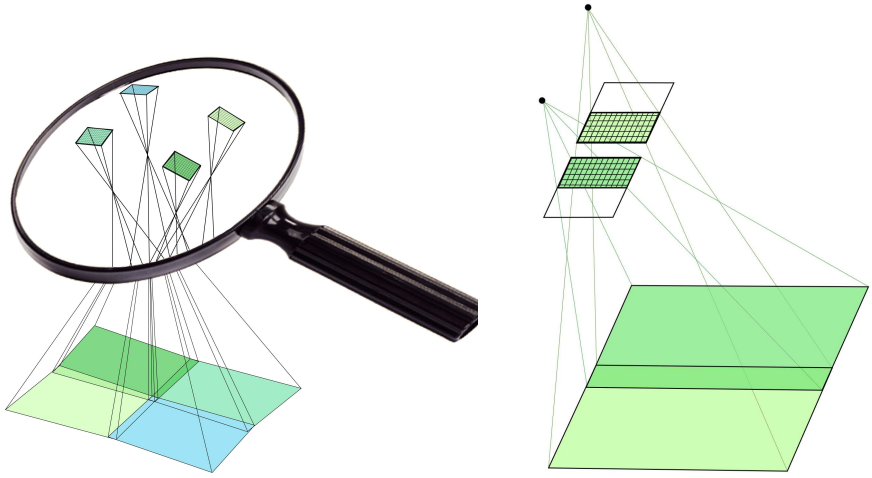
1.2.2 Multiple head

The most intuitive way to design a digital mapping camera would be to replace the former analogue film by a 2-D electronic sensor element or sensor matrix. Indeed, such approach was pursued in consumer-grade photography. Unfortunately, the size of a CCD frames is physically limited by the supporting electronics. Therefore, it took some years until a special multiple-head concept was developed based on cluster of CCD sensors with format comparable to the former 35 mm format (24 mm x 36 mm negative). Such a design employs several individual camera heads, each one equipped with one or more CCD frame sensors that are all firmly attached to one airborne platform. Due to the special geometrical arrangements the individual CCDs of a smaller format connected to separate camera heads are generating multiple smaller format images with certain overlaps.This allows for the generation of one synthetic large format image afterwards, which is obtained by re-sampling the individual smaller format single images to a virtual large format on one focal plane. In other words, the virtual large-format image can be used in later production in the same way as any other frame images. The only difference is its derivation from a virtual camera instead of a physically existing camera.

Often multi-head frame cameras are designed in a way that the individual heads (generally two or four) are arranged with slightly oblique viewing directions. Such inclined installation of camera heads results in four overlapping images, so-called butterfly pattern, which is necessary to form a virtual image of a large format (see Fig. 1.2). These overlaps are necessary for the later transfer of corresponding points (also called tie-points) or tie-features that enables the merging of several smaller images into one virtual image of a larger format.

Different to the concept of tilted camera heads that generate overlapping images, some installations rely on nadir looking camera heads only. One approach is to slightly shift the CCD frames against each other in the neighboring camera cones. There the CCDs are not placed in the center of

each focal plane but slightly de-centered, shifted to the direction of opposite edges of the individual focal planes. As the images are taken at the same time part of the covered scene overlaps and therefore the images can be merged together (see Fig. 1.2).



(a) "Butterfly arrangement" with 4 tilted heads, each head with one CCD located in center position.

(b) "Shifted CCD arrangement", here with 2 nadir looking heads, each with one CCD located in shifted position.

Figure 1.2: Camera concepts based on CCD frame arrangements, synchronous recording, courtesy ifp-Stuttgart.

1.2.3 Syntopic frame

In the so far presented concepts of multi-head frame cameras the images originating from the individual cones are taken at the same time, i.e. the image exposure of individual camera heads is synchronized. A different concept relying on multiple nadir-looking camera takes images at different times, however, over the same place. Such approach is called syntopic image recording and it is based on the idea that multiple camera cones are arranged in a line, which coincides with the main flight direction (Fig. 1.3). If the different camera heads take their images one after the other and this time shift exactly corresponds to the velocity of the camera movement the images will be taken at the same position. As a result the camera stations for all images are the same. Different to the previously described system layout, the camera cones contain between one to four CCD frames which are installed in different arrangements in their focal planes. Dependent on the individual arrangement of CCD frames within the different camera heads, overlapping images are generated in object space which again can be merged together afterwards. The concept of image formation from syntopic imaging

is illustrated in Fig. 1.3. As can be seen on the figure, up to four CCDs are placed within one single focal plane in a special pattern. The cone containing the four CCD-sensors in its corners defines the virtual frame obtained after image stitching. It is named the primary or master cone. The remaining cones are used to fill the gaps in between.

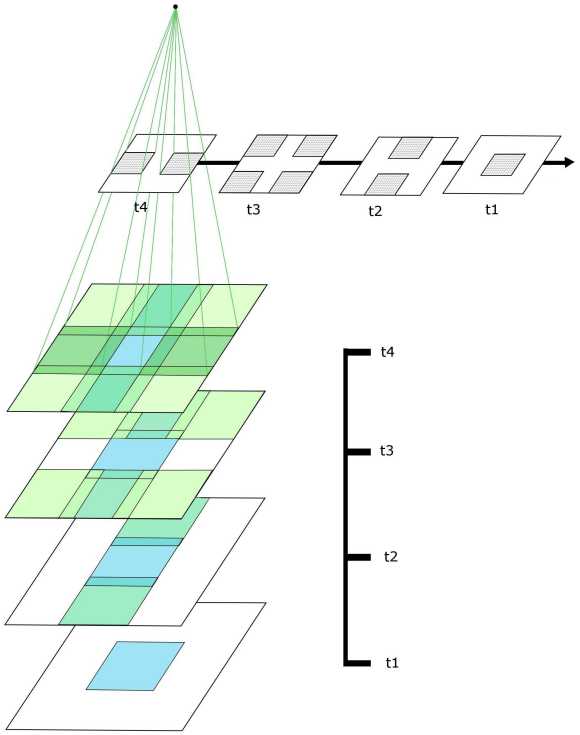


Figure 1.3: The syntopic imaging concepts, courtesy ifp-Stuttgart.

1.2.4 Comparison of concepts

Fig. 1.4 compares the two ground patterns of the two main multi-head concepts: four camera heads are used for both installations, but in the first case the images are taken at the same time (synchronously) with tilted camera heads. The Fig. 1.4-left shows the particular footprint of such setup in the object-space, where the different colors indicate each of the four camera heads. Due to their off-nadir viewing the four images have individual perspective displacements. This tilt influence the imaging of same objects in two camera heads, which is especially of concern in the overlapping parts. The effect is dependent on the height differences in object space but should be negligible in most application scenarios (Dorstel et al (2003)).

The syntopic image recording delivers a different pattern (see Fig. 1.4-right). Again, the color shades indicate the arrangement of CCD frames in

the four participating camera cones. All cameras are recorded at the same place (due to the small time interval between the different recordings) in the nadir looking direction. Therefore, they should have the same perspective displacements as long as the difference in the perspective center coordinates or off-nadir variations are negligible. Again, the overlapping regions between them are used for the formation of large-format imagery.

The dashed frame in Fig. 1.4 indicates those parts of the images which are used to form the virtual image of a large format. As can be seen, smaller parts at the corners of the butterfly pattern are lost. This is because the format of the virtual image is chosen to be rectangular. In case of the syntopic imaging, the virtual image may use the full part of the individual frames. Practically, a small margin is also cut-off in this approach.

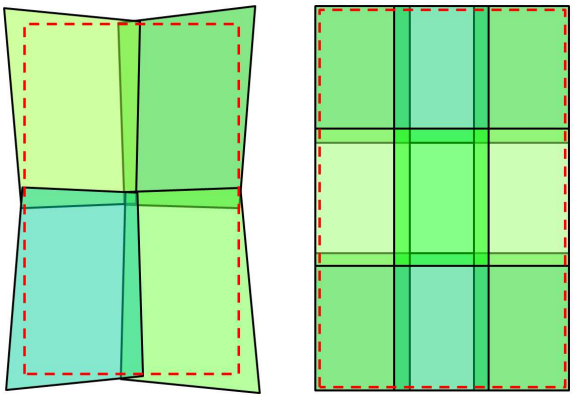


Figure 1.4: Multi-head concept to generate large format frames: Synchronous imaging using 4 tilted camera heads (left) and syntopic imaging using 4 nadir pointing camera heads (right), courtesy ifp-Stuttgart.

1.2.5 Virtual frame

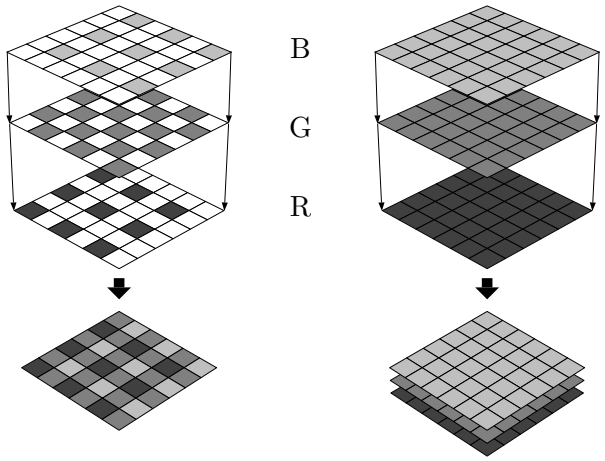
The multi-head concepts allow for the generation of virtual images of a larger format. For this purpose several individual images are re-sampled to a previously defined virtual-focal plane. This is based on the individual interior orientation of each camera head and their orientations relatively to each other. The process is called inter-cone orientation or *image stitching*. The knowledge of interior orientation (see Sec. 4 for definition) of every contributing camera head is necessary, to exactly reconstruct the 3-D image rays originated from each camera-head pixel. Moreover, the relative orientation between the different camera heads (represented by 6 independent parameters) is required to determine their relation to the virtual focal plane. All together this defines the correct position where the image ray intersects the virtual large-format plane. The interior orientation of the camera heads is assumed to be known, whereas the orientation between the different cam-

era modules is derived from conjugate points measured from the overlaps between the different images (Dorstel et al (2003); Ladstadter et al (2010)). Even in multi-head cameras, where the different cones are mounted on one platform and images are taken in synchronized mode, the existence of such overlapping regions is necessary to control the stability of the orientation between the individual camera cones.

1.2.6 Color generation

The generation of multi-spectral images from frame-based sensors can use several concepts. Many digital frame cameras, especially those designed for the consumer market, use the so-called Bayer-pattern approach, where typically red, green and blue (RGB) filters are arranged over every pixel on the CCD sensor in a special pattern. Thus each pixel become sensitive only to one of the three base colors. The color is then derived through interpolation from neighboring pixels that contains the RGB components.

An alternative concept is employing separate camera heads for each of the requested multi-spectral channels. Appropriate filters let each CCD array only capture the corresponding color information. Red, green, blue and additional near infrared spectral bands are most common. Full RGB is derived through a so-called registration of color bands. The different images are overlaid to generate full color after adding three selected color. In order to guarantee congruent features in the different color images a geometrical (2-D) transformation of images based on corresponding matched points between the different channels is necessary.



(a) single CCD using Bayer pattern

(b) individual CCDs, one for each color channel

Figure 1.5: The two concepts to obtain colored RGB digital images from frame sensors.

Fig. 1.5 illustrates both concepts. When using the Bayer pattern approach one CCD frame is sufficient to capture the full color information, but due to the pixel-wise color filters, each pixel only contains the color information of the corresponding channel. Additionally, 50% of all pixels are sensitive to the green spectral band while only 25% are sensitive to red and blue respectively. This is done to adapt to the color sensitivity of human eyes. Full color information is then derived from color interpolation of the neighboring pixels. If alternatively several CCD frames are available with each of those being sensitive for one color, the radiometric information can be delivered for each pixel with same resolution. Each pixel on the ground is imaged in all three color bands. Still the three individual images have to be merged before the full color RGB image is derived. This approach typically demands at least one camera heads per color band or alternatively a beam splitter in the camera optic to separate the different color bands within one single camera head.

1.2.7 Color resolution

Multiple head cameras usually separate registration of panchromatic (grey-values) and color channels. As color is often generated using Bayern pattern, this results into somewhat lower geometric resolution with respect to the panchromatic (PAN) image. Also, the design of multi-spectral channels is likely to set even lower spatial resolution compared to the large format virtual pan image. High-resolution color imagery is than obtained from post-processing, where the lower-resolution color channels are combined with the high-resolution PAN images. This process is named pan-sharpening and is frequently used in satellite imaging. The ratio between the spatial resolution of the pan and color channels is called pan-sharpening ratio. Different approaches are used for pan-sharpening Gonzalez and Woods (1992). The methods can be classified in substitution approaches, arithmetic and filter based techniques. The preservation of original radiometric color information, depending on the algorithm is exemplarily discussed in Ehlers et al (2010). The basic idea of the pan-sharpening concept is illustrated in the Fig. 1.6. This example is taken from digital airborne image data. After the fusion of the lower resolution RGB image (b) with the higher resolution PAN image (a), pan-sharpening delivers a color image with higher geometric resolution of the PAN channel (c). Within this example one low-resolution RGB pixel corresponds to 4 x 4 high-resolution PAN pixels, which equals to a 1:4 PAN-sharpening ratio.

Figure 1.6: The concept of pan sharpening to increase the spatial resolution of color imagery. (a) High-resolution pan, (b) RGB low resolution, (c) RGB after pan sharpening, courtesy ifp-Stuttgart.

1.3 Line sensors

1.3.1 Concept

The previously described group of digital mapping cameras was based on the CCD (or CMOS)-frame concept. Digital imaging from moving platforms might also be based on single or multiple CCD-lines. Similarly to an office scanner, only one or few CCD lines is arranged perpendicular to the principal moving direction of the sensor. A full 2-D image is indirectly obtained due to the sensor's motion. While the platform is moving, the two dimensional image data are captured, with the CCD line(s) almost continuously recording. This line scanner concept is also named pushbroom scanning. Digital pushbroom scanners were first introduced into satellite imaging, later also to airborne image acquisition. The principal advantages of a pushbroom scanner is the possibility of extending the length of CCD lines beyond the limits of frame-sensors and thus obtaining larger swath and ground coverage. In the modern airborne imaging this advantage is challenged by the introduction of previously discussed virtual frames.

If only one CCD line is used than the line-image has an extension of just one pixel in flight direction. Such a line-image is acquired at one distinct point of location and time. The image width equals to the number of pixels per line, i.e. the length of the CCD line. The consecutively imaged lines form the *image strip*, which also is named *image scene*. Notice, that each individual line-image has its own exterior orientation elements, i.e. position and attitude. This is relevant for the later orientation process of the push-

broom image data (Sec. 4). The obtained pixel size on the ground depends on the sampling time of the system and the speed of the platform. Since the linear sizes of the ground pixel in- and across-track are independent, quadratic pixel on the ground are only obtained if the so-called pushbroom condition is fulfilled. The corresponding ground sampling distance (GSD) is derived from the relation:

$$\begin{aligned} \mathrm{GSD}_{\mathrm{along}} &= v \cdot \Delta t \\ \mathrm{GSD}_{\mathrm{across}} &= \Delta y \cdot m_b = \Delta y \cdot \frac{h_g}{f} \end{aligned} \tag{1.1}$$

where Δt is the sampling time, v the speed of the platform, h_g the flying height above ground, f the camera focal length, m_b the image scale, Δy the pixel size across flight direction.

Typically, more than one CCD line is used in a line scanner system (Fig. 1.7). If two or more CCDs are arranged in one focal plane, along-track stereo viewing becomes possible, where the desired stereo angle is constant and exactly defined through the distance between the different lines in the focal plane. Multiple CCD lines are also necessary to record different color channels. Different to the frame-based approaches no additional sensor heads are necessary for color and multi-spectral imaging. Additional lines are simply placed in the same focal plane that is already used for the panchromatic channels. Often at least three pan-chromatic channels as well as four multi-spectral channels are used. All CCD lines provide the same number of pixels. Thus pan-chromatic and multi-spectral images are obtained with the same geometric resolution, which is again a characteristic different from the frame based sensors. Even though almost all systems have more than three lines, such pushbroom systems often are referred to as three-line scanners. This is named after the three pan-chromatic lines.

1.3.2 Geometrical configurations

Since the physical location of each of the lines on the focal plane is different, each CCD line provides a different viewing direction, which allows multiple stereo angles within one flight line. In the Fig. 1.7 multi-spectral channels are exemplarily placed in the nadir viewing direction plus the three additional panchromatic lines in forward, backward and nadir direction. Thus three different stereo angles are possible, namely between the forward and nadir, backward and nadir, and forward and backward view. The color lines might also be arranged in off-nadir direction. In some systems more than one linear CCD is used for each color channel. If those are placed at different positions in the focal plane this also allows color and multi-spectrum stereo viewing capabilities.

As mentioned the stereo capability depends on the different viewing directions due to the parallax effect. Nevertheless, such parallaxes also appear

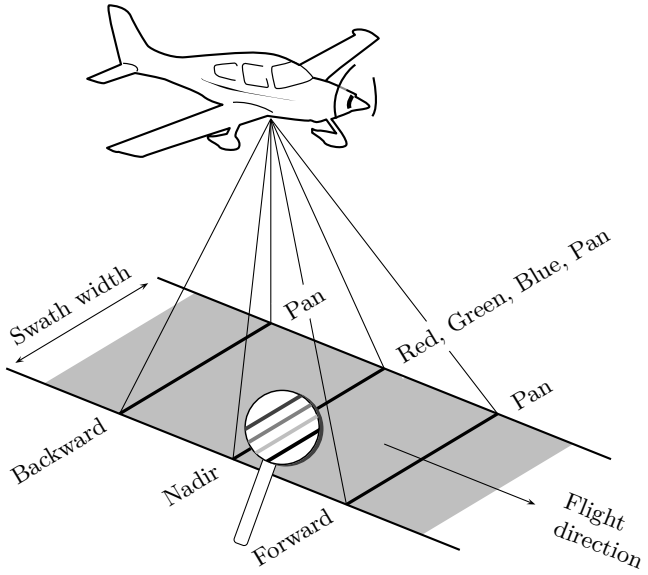


Figure 1.7: The concept of an airborne line scanning system, after Petri (2000).

in the different multi spectrum lines which are needed to later obtain the full color images by combination of individual spectral bands. Even though these color/multi-spectral lines typically are mounted as close as possible their displacement will cause different perspective distortions in each spectral band. The larger the distance between the different spectral CCD lines is, the larger the influence of these displacements is. In order to correct for these effects two options are possible:

- The first is that the full color image is always generated in the orthophoto domain (Sec. 5.2). The orthophoto processing corrects for any displacements in the perspective images, also considering the influence of height variations of the imaged scene. If each color band image is fully rectified the individual bands can easily be overlaid to obtain the full color image.

- Alternatively, this problem can be overcome if so-called beam splitters are installed in the optical system of the pushbroom scanner. Such installation allows to exactly co-register the four different color bands (see Fig. 1.8). With that each of the color bands has the same perspective geometry. Such beam splitters are located in the optic module of the camera, between the lens and the CCDs.

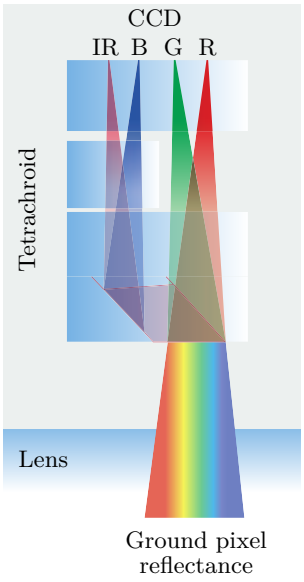


Figure 1.8: A beam splitter used in a push-broom sensor, (© Leica Geosystems).

1.3.3 Image staggering

The number of pixels per line directly defines the obtainable maximum swath-width of the system. If for example a GSD of 10 *cm* is requested, the resulting swath will be 1200 m if the system is based on 12000 pixels per line CCDs. A larger strip width improves the efficiency of the data capture as this influences the number of strips to be flown to image a project area.

The width of the swath can be further extended if the image lines are staggered. The staggering means that two CCD lines, so-called A and B lines, are fixed at almost the same position on the focal plane, but shifted by half a pixel in the across-track direction. Fig. 1.9 shows the arrangement of a staggered line with 6.5 x 6.5 μm^2 pixel size. Here the distance between the two lines equals to 4 pixels. While both lines are imaging the same scene, their respective pixel centers are shifted by half a pixel. This obviously increases the sampling interval across flight direction by a factor of two. The line frequency, i.e. the sampling rate in flight direction, is then adapted according to the new sampling rate in the direction of the CCD line. The A and B lines take one image each which can be superimposed and combined to a new image with a (nominal) doubled resolution compared to the original images.

Fig. 1.10 illustrates the concept of a staggered array. Here two CCD lines with only 3 pixels per line are combined. It can be seen, that when employing the staggered mode, each line acquires pixels of rectangular shape.

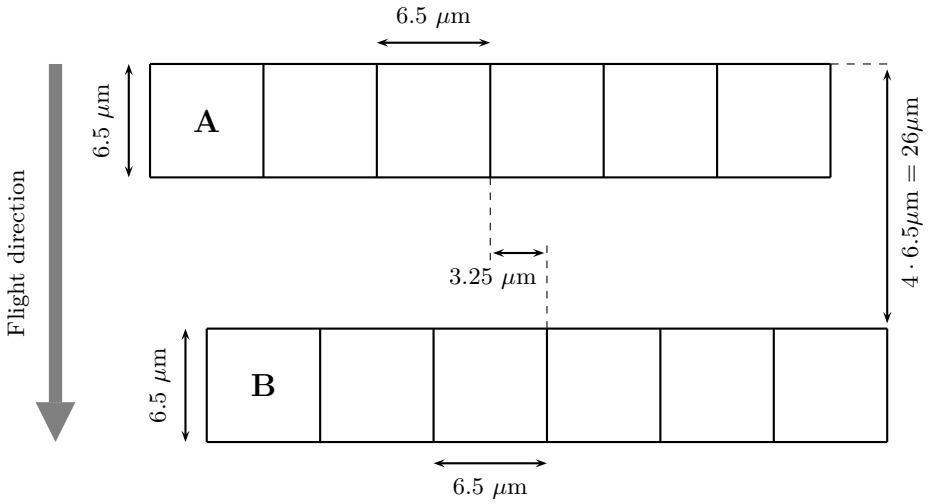


Figure 1.9: Staggered line arrangement, situation in focal plane.

The sampling rate in the flight direction is duplicated, in order to prepare for the later staggering where the form of staggered-pixels becomes square. Due to the small distance between the two lines, slightly different parts of the flown area are imaged at the same time. In this example line A at time t_2 covers the same area which already was imaged in line B at time t_0. Due to the half a pixel shift between line A and B in the focal plane the two sampling patterns of both scenes on the ground ideally complement each other, which will deliver a combined product with increased resolution. The figure also shows, that the refinement of resolution fails, if the requested ideal sampling pattern overlap is not done correctly. As the lines A and B are physically shifted, there is a small time difference between their exposures to capture the same object on the ground. Therefore, the staggering is affected by the relative change in the sensor attitude during data acquisition and hence the quality of platform stabilization. This effect is less critical for satellite-born sensors where the trajectory is much smoother as compared to airborne platforms (Petri and Walker (2007)).

Another approach to increase resolution and swath-width is the employment of multiple CCD lines which are shifted against each other across flight direction. Since these CCD lines cannot be stitched together directly, a stepped arrangement is necessary (Fig. 1.11). Although this approach was so far adopted only in the early stages of line-camera development, it is used for obtaining high-resolution satellite imagery. Since the focal plane layout is complex in this configuration, additional processing is required to overcome the discontinuities and misalignment between the lines.

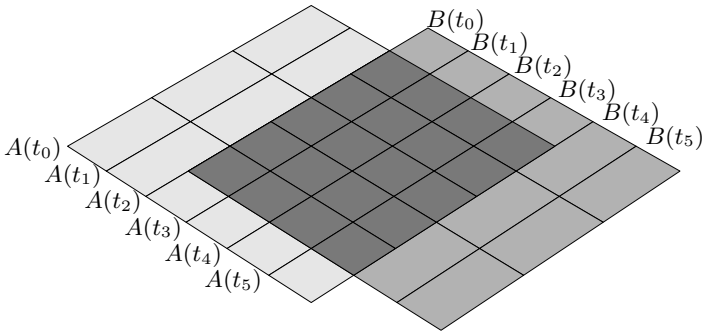


Figure 1.10: Sampling pattern on the ground with the concept of line-staggering, courtesy ifp-Stuttgart.

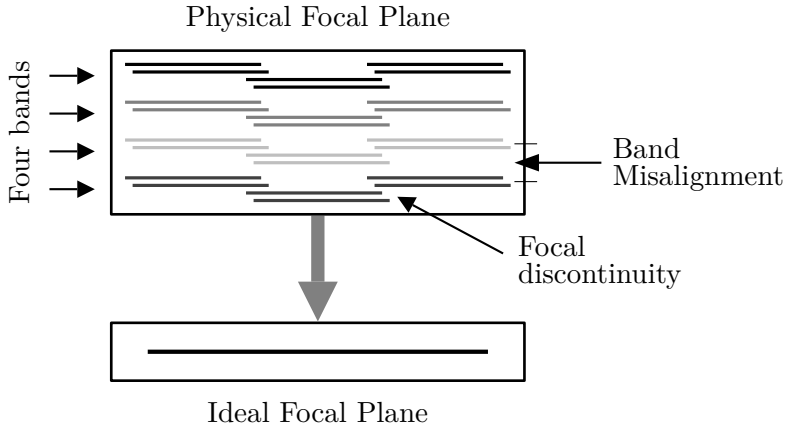


Figure 1.11: Stepped arrangement of multiple CCD lines.

1.4 Lidar

The acquisition of terrain structure is very efficiently achieved by optical sensors such as Radar and Lidar. Both methods are actives, i.e. an energy is emitted from the sensors and the its reflection by the object is recorded and processed. The terrain models of highest precision and resolution are usually obtained by Airborne Laser Scanning (ALS), which is in the primary focus of this section.

1.4.1 Laser ranging

Introduced towards the end of the last millennium, Lidar is is one of the most important geospatial data acquisition technologies. Together with the state-of-the art navigation technology mobile Lidar systems are capable to collect three dimensional data in large volumes, high density and at unprecedented accuracy.

The fundamental principle of laser-ranging is the ability to measure the travel time t of an emitted laser pulse along its path from the instrument to the target and back (Fig.1.12). Hence, the distance ρ from the ranging unit towards the target is deduced by the following relation:

$$\rho = \frac{1}{2} c\, t \tag{1.2}$$

where c is the speed of light. As shown in Fig. 1.12 the laser-ranging unit comprises an emitting laser and an electro-optical receiver. The transmitting and receiving apertures are oriented in the same direction, to ensure that the system will detect the target the transmitter points to. The size of the laser footprint is a function of the distance to the target and the divergence ϵ of the beam. The angle ϵ defines the instantaneous field of view (IFOV). The IFOV usually spans from 0.1 to 3 mrad.

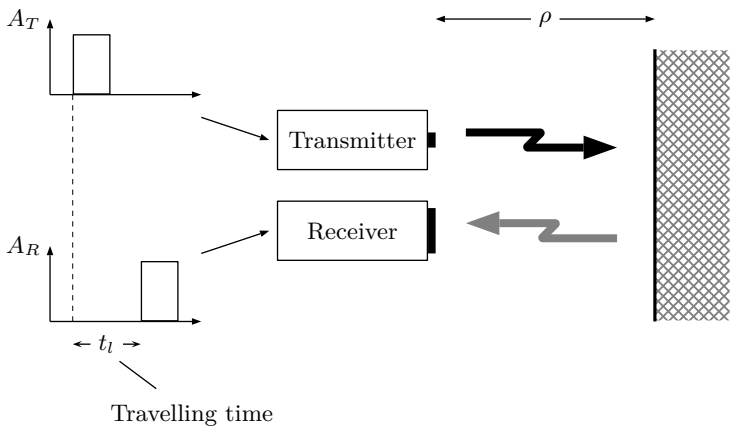


Figure 1.12: Lidar ranging principle. A_t and A_r are the amplitudes or intensity of the transmitted and received pulses, respectively, after Wehr and Lohr (1999).

There are two technological principles of laser ranging that are implement in mapping applications: the continuous wave (CW) lasers and the pulse lasers. In CW lasers the radiation is emitted as continuous beam instead of sequence of discrete pulses. This limits the power of the CW laser to terrestrial laser scanning, although there are exceptions (Hug, 1994). CW lasers deduce the range by comparing the phases between the outgoing and incoming signals. The phase difference of the received light wave is proportional to the travel time of one wavelength (period) and thus to the range:

$$t = \frac{\phi}{2\pi} T + nT \tag{1.3}$$

where t is the total elapsed time, ϕ the phase difference of the returned wave, T the period of the modulated signal and n the number of full wavelengths included in the distance from the transmitter to the receiver. As the phase information is ambiguous for a single measurement, the CW instrument needs to employ a way for its resolution. This can be achieved by various means, most often by modulation of frequency or by following range changes (Wehr and Lohr, 1999).

Although CW lasers reach higher ranging accuracies, today's airborne Lidar systems almost solely use pulsed ranging. A pulse laser functions according to the Eq. 1.12. The wide spread of pulse lasers is due to two technological advances. First, the progress in accurate quartz-stabilized oscillators enables determining the elapsed time between the emission and the reception at picosecond (ps) level (i.e 10^{-12} s); second, the existence of powerful laser sources with fast shutter limits pulse duration below nanosecond (ns) (i.e.$< 10^{-9}s$) level. Today's pulse lasers achieve cm to mm-level ranging resolution in long and close-range instruments, respectively (Lohr et al, 2010). In long-range (airborne) applications the different implementations of pulse-based range-finders can be distinguished:

- *Linear mode - discrete echo*: After emission of high-energy, longer laser pulse, a representative trigger signal of a return (an echo) is detected in real-time using analog signal processing. As a discrete pulse is spread in space along its line of sight, part of its energy can be reflected by multiple targets. This allows to scan even through the canopy, because the spacing between the leaves and branches allows parts of the pulse to penetrate further to the ground, while some energy is reflected immediately. This principle is schematically depicted in Fig. 1.13. As shown in the third plot of Fig. 1.14, the partial reflections are detectable above certain threshold as distinct peaks in the gathered return signal. These are then discretized into separate echoes. Systems based on this principle can record several returns with minimum separation between successive pulses of several decimeters.

- *Linear mode - full-waveform*: Employing also high-energy, longer laser pulse, these instruments digitize the entire analogue echo waveform, i.e. the time-dependent variation of received signal power, for each emitted laser pulse (lowest plot in Fig. 1.14). This approach overcomes the pulse- separation limit present in discrete echo systems and allows finer resolution in the range. The digitization is performed typically on several channels with an interval of 1 ns, which corresponds to spatial quantization of about 0.15 m. The determination of the individual echoes is usually performed after the mission, although modern airborne laser scanning (ALS) systems perform full echo digitization and waveform analysis in real-time.

- *Geiger-mode*: These devices emit medium energy of short laser pulses into one beam of certain opening. The detector side contains several thousands of pixels that are sensitive to a weak return (few photons) in a binary manner. This so called "Geiger" counters requires hundred times less power to register a return than the linear mode Lidar detectors. The large sensitivity of Geiger detectors allows considerably longer ranging than for linear-mode scanners. Coupling long-ranging capability with the employment of large number of small detectors and high repetition rate (hundreds of MHz) allows maintaining few pulses per m^2 from 5-10 km above ground, which increases considerably the swath width and thus the productivity. However, the first generation of these detectors allow to register one (first) echo only with considerably lower precision than that of linear mode scanners. These instruments are yet to be introduced into civilian airborne laser scanning.

- *Single-photon*: These devices emit very low energy and short laser pulses into approximately hundreds of beams. There are separate detectors per beam containing on the order of hundreds of pixels. Each pixel can detect single photon return at high resolution (<0.1 m) while registering multiple returns per laser shot with a separation of 1-3 ns. As the system is able to record multiple event per pixel channel and per laser shot in one beam while employing multiple beams several million points per second are scanned with multiple stops. This technology is therefore even more productive that Geiger-mode scanners, albeit not yet as precise as linear-mode lasers. First commercially available ALS of this type was introduced in 2018. At the same year a single-photon scanner was placed on an orbit of a satellite mapping ice (ICESat-2).

Most commercial laser rangers operate between 900 and 1500 nm (near-infrared) wave length, while single-photon lasers currently use 530 nm (green laser). The amplitude of the backscattered energy A_r is in practice referred to as intensity and is recorded together with the distance observation. (This reference is common but incorrect due to adaptive amplification of the received signal according to its long term average.) Its value depends on several factors:

- *Laser wavelength and target reflectance*: Varying the laser wavelength results in different reflectance responses on the same surface. For example, a laser using wavelength 1500 nm has good reflectance responses on dark surfaces and man-made structures, whereas surfaces with water content (i.e. glaciers, snow) reflect weakly. On the other side, systems with shorter wavelength ($<$ 1000 nm) have good reflectance on snow cover but are less optimal for mapping in urban areas. At the same time, objects with high reflectivity such as street

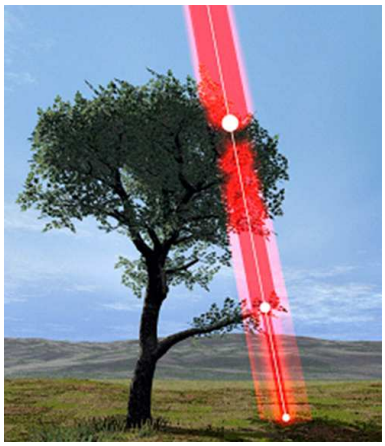


Figure 1.13: Principle of multiple echoes from in Lidar, after Schaer (2009).

mark paintings or cement contrast distinctly with objects of low reflectivity such as coal or soil.

- *Incidence angle of laser beam*: The level of the backscattered signal is a function of the integrated energy distribution across the whole footprint. Accordingly, the larger the incidence angle, the larger the footprint and consequently the smaller the backscattered energy.

- *Atmospheric illumination and attenuation*: External illumination, such as sun light or reflectance from clouds acts as noise in the returned signal. Additionally, light propagation in the troposphere is affected by both, scattering and absorption characteristics of the atmospheric medium, thus reducing the reflected energy.

1.4.2 Profilers

Laser profilers measure the distances to a series of closely spaced points distributed along a line on a terrain. In space or airborne applications the profiler is a simple laser ranger (often called laser altimeter) that is pointed towards the ground. Such altimeter measures the distances while is moved over the ground on board of a vehicle. As schematically shown in Fig. 1.15 the 2D terrain profile is obtained when the altimetric distances are connected to the position and orientation of the laser profiler. Before the invention of satellite positioning, the precise measurement of carrier's position was difficult to achieve, reason for which the laser altimetry was used almost exclusively on space-borne platforms. There, the motion was determined by satellite-tracker observations and by appropriate modeling of the trajectories. This laser technology was first used to determine sea surface topography, ice cover, desert topography, etc. (e.g. TOPEX/Poseidon,

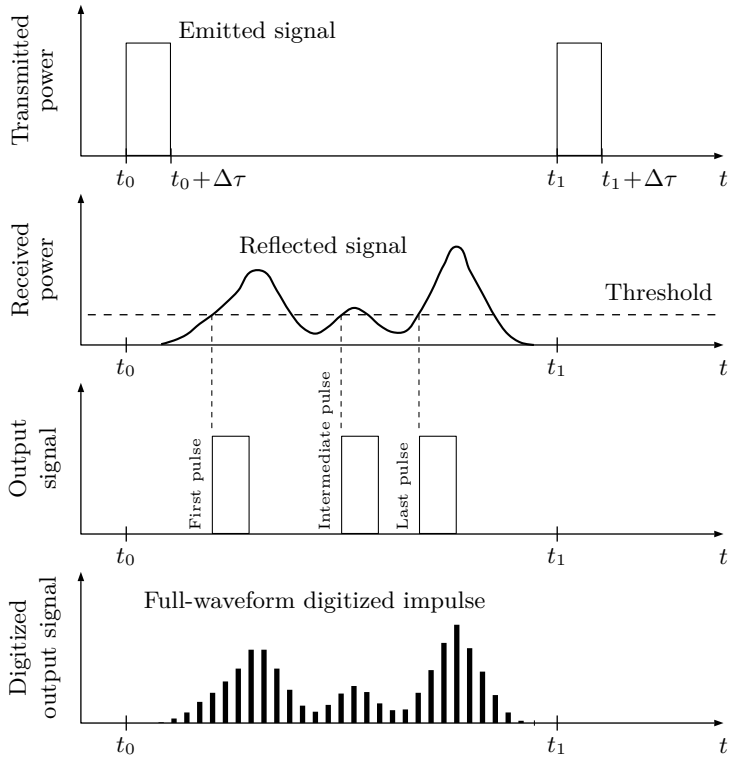


Figure 1.14: Emitted and received impulse for discrete echo scanners and full-waveform scanners, after Schaer (2009), courtesy author.

Jason-1, Envisat satellite missions). Later, more sophisticated laser instrumentation allowed the conjoint observation of the Earth surface relief and vegetation canopies (Shuttle Laser Altimeter (SLA), Carabajal et al (1999)) or distribution of clouds and aerosol (Geoscience Laser Altimeter (GLAS)).

Airborne laser profilers are less common than laser scanners. Nevertheless, these instruments are still used for surveying slowly changing surfaces such as ice-covered terrain (Spikes et al, 1999), lakes or costal water bodies. The latter applications are often connected to the calibration of satellite altimeters or to the study of local gravity field (Geiger et al, 2009).

In a terrestrial or ground-based laser-profiler, a sequence of distance measurements is executed in a series of steps with the slight change of laser-beam orientation between them. Thus, the 2D elevation profile Δh with

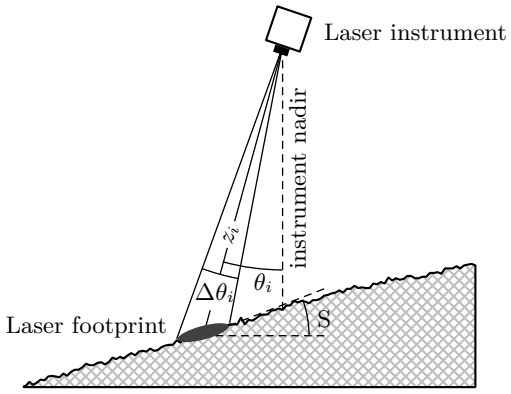


Figure 1.15: Lidar profiling from a space-borne platform using a laser altimeter.

respect to the leveled instrument is obtained as

$$\Delta h = \rho \sin(\theta) \tag{1.4}$$

where ρ is the slant distance and the θ the recorded vertical angle. This results in a two-dimensional profile or vertical cross section of the ground. The terrestrial laser profiler is essentially a 2D laser scanner which is described in the following section (Sec. 1.4.3).

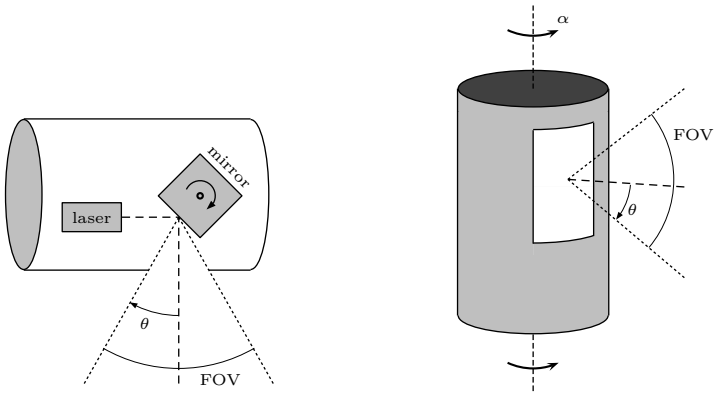
1.4.3 Scanners

Lasers scanners combine a laser range-finder with a scanning mechanism (e.g. a mirror) to direct the laser beam into desired direction. The scanning mechanism has either one or two degrees of freedoms that are used to create 2D or 3D profiles, respectively. Frequently, the 2D scanning mechanism is used (Fig. 1.16a), from which the 3D profile is created by either

- rotating the whole scanner assembly along a vertical axis, as would be the case in static Terrestrial Laser Scanning (TLS) (Fig. 1.16b)
- movement of the carrier in kinematic laser scanning (airborne or vehicle-based scanning)

Thus, in the latter case, the motion of the platform enables along-track scanning, while the mirror deflection provides across-track scanning. The total across-track scanning angle defines the swath width or scanner's field of view (FOV). The swath width SW on the ground can therefore be computed as a function of the flying height h and the instrument's FOV ϕ_{max} as

$$SW = 2H \tan \frac{\phi_{max}}{2} \tag{1.5}$$



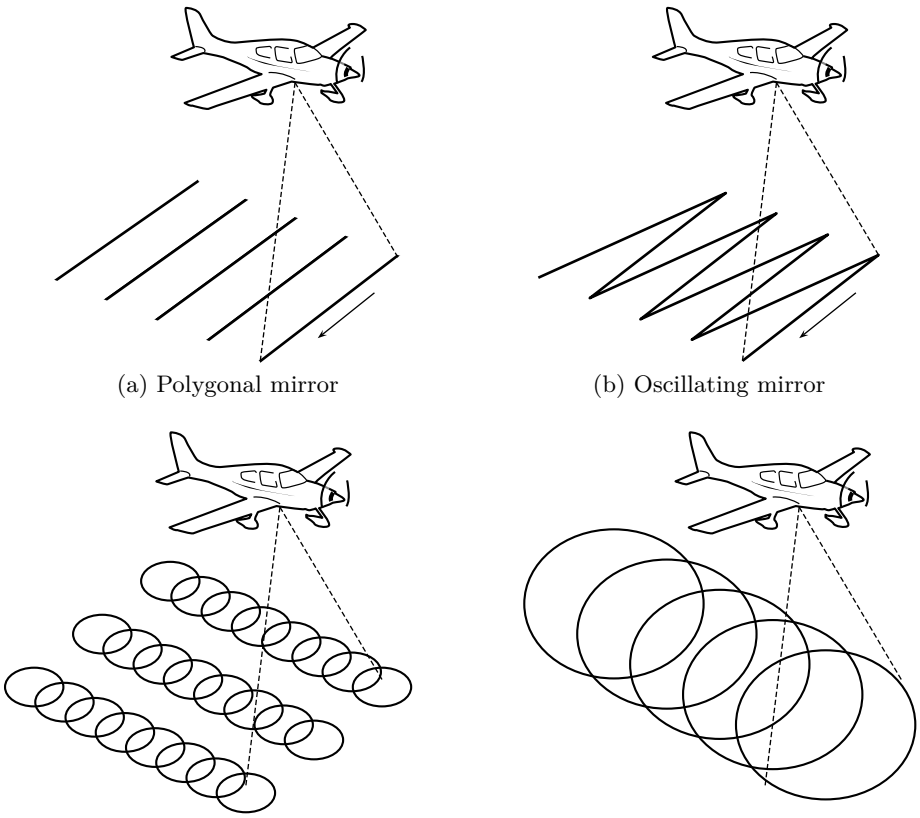
(a) 2-D laser-scanner with a rotating polygonal mirror

(b) 3-D scanner based on the same principle

Figure 1.16: Lidar scanners.

The typical FOV of today's scanners is 50-60° in airborne and 80-180° in terrestrial scanning. Several scanning mechanisms exist. The principle of several scanning principles used on airborne platforms is depicted in Fig. 1.17 and their comparison is provided in Table 1.1.

The potential of employing laser ranging for navigation and collision avoidance systems initiated the development of devices operating over shorter distances (<100 m) without the scanning mechanism. There, few tens of lasers are arranged in a line-array with a regular angular separation and FOV of 30°. The 3-D profile is created by rotating the whole assembly, similary as in (Fig. 1.16b), nevertheless with rotation rates up to several tens of Hz, resulting in high data collection rate. Although the ranging is generally less precise than for scanning lasers, after proper calibration (Glennie and Lichti (2010); Glennie et al (2016)), these devices have applicability in mapping from ground vehicles and UAVs.



(a) Polygonal mirror (b) Oscillating mirror

(c) Fibre scan (d) Palmer scan

Figure 1.17: Different scanning patterns (after Morin (2002)).

Scanner frames

The definition of a scanner frame is chosen arbitrary and therefore differs among manufactures. The following definition applies to several systems and can be ported to other instruments by simple permutation of axes. The location of a point within in scan line j can be conveniently expressed either by polar or Cartesian coordinates, with the former is usually used. Considering the situation as depicted in Fig. 1.18, the relations between the *range measurement*$-\rho$, the encoder *horizontal angle*$-\theta$ and the *vertical angle* $-\alpha$ with respect to scanner frame defined in Cartesian coordinates are:

$$\rho_{ij} = \sqrt{x_{ij}^2 + y_{ij}^2 + z_{ij}^2} \tag{1.6}$$

$$\theta_{ij} = \arctan\left(\frac{y_{ij}}{x_{ij}}\right) \tag{1.7}$$

$$\alpha_{ij} = \arctan\left(\frac{z_{ij}}{\sqrt{x_{ij}^2 + y_{ij}^2}}\right) \tag{1.8}$$

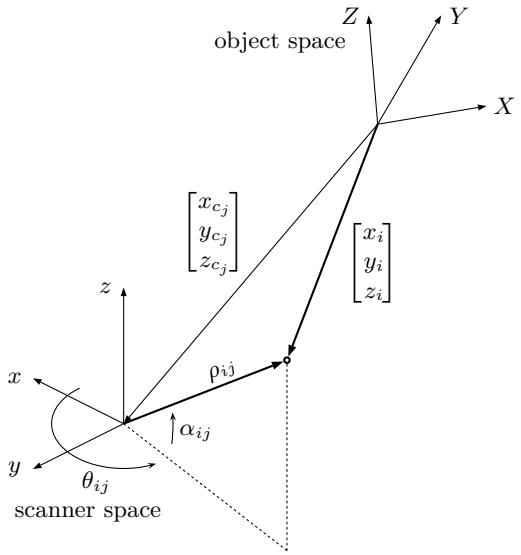


Figure 1.18: Scanner frame and observation geometry.

In case of a 2D scanner (e.g. airborne or terrestrial mobile scanning), the angle α is zero and the Cartesian coordinates of the target are expressed as

$$\mathbf{x}^s = \rho \cdot \begin{pmatrix} 0 \\ \sin\theta \\ \cos\theta \end{pmatrix} . \tag{1.9}$$

Mechanism	Characteristics
Polygonal mirror	⊕ Constant rotation avoids mirror distortions due to additional force ⊕ Provides regularly spaced sampling along and cross-track ⊖ Observations can be taken only at small portion of each mirror facet ⊖ FOV is fixed and cannot be adapted ⊖ Systems are limited to lower flying heights above ground ($< 1000\,\mathrm{m}$)
Oscillating mirror	⊕ Continuous data acquisition possible as mirror points always towards ground ⊕ Possibility to compensate aircraft rotation around roll ⊕ FOV can be adjusted ⊖ Mirror acceleration causes systematic distortions due to tortion ⊖ Z-shaped irregular sampling with lower density at nadir
Fibre scan	⊕ High scan rate possible due to fewer and smaller moving parts ⊕ Scan rate sufficiently high to provide along-track overlap ⊕ Regular ground sampling ⊖ FOV is limited ⊖ Across-track spacing is fixed
Palmer scan	⊕ Scanning is performed twice, each time from a slightly different perspective ⊕ Scan rate sufficiently high to provide along-track overlap ⊖ Increased complexity of two mirror motion is harder to calibrate and encode ⊖ FOV is limited ⊖ Across-track spacing is fixed
Line array	⊕ Faster than a scan thanks to concurrent use of many lasers ⊕ 3-D scan is created / updated rapidly ⊖ Limited to close-ranging with lower accuracy ⊖ FOV is limited ⊖ Across-track spacing is limited to the number of lasers

Table 1.1: Comparison of different scanning patterns used in mobile laser scanning: ⊕ advantages, ⊖ disadvantages.

Chapter 2
Navigation Sensors

2.1 Mapping prerequsits

Spatial interpretation of remotely sensed data requires determination of the geometric relation between the sensor and the real world. Once these relations are found, the data can be interpreted in some reference frame (local or global). In literature this process is referred to as *georeferencing*, *geocoding* or *(sensor) orientation* and concerns the following components (Fig. 2.1):

- The determination of internal geometry of the sensor (*interior orientation*).
- The determination of sensor orientation relatively between scenes (*relative orientation*) or with respect to an external frame (*absolute orientation*).

According to the sensor type the exterior orientation parameters (EO) may include position, attitude (e.g., cameras, scanners) and velocity (e.g., RADAR). For passive sensors (e.g., frame or line cameras), these parameters may be deduced indirectly from data overlaps and ground control features distributed across the scene (*indirect* sensor orientation), by determining them with a suitable navigation system (*direct* sensor orientation), or by combining both approaches (integrated sensor orientation, Sec. 4). Active sensors (e.g. laser, radar), on the other hand, urge the use of direct sensor orientation. Due to the sequential measurement principle and the motion of the carrier vehicle in mobile mapping the EO parameters differer for every object point. The following text provides an overview of the navigation technology that facilitates tremendously the problem of sensor orientation.

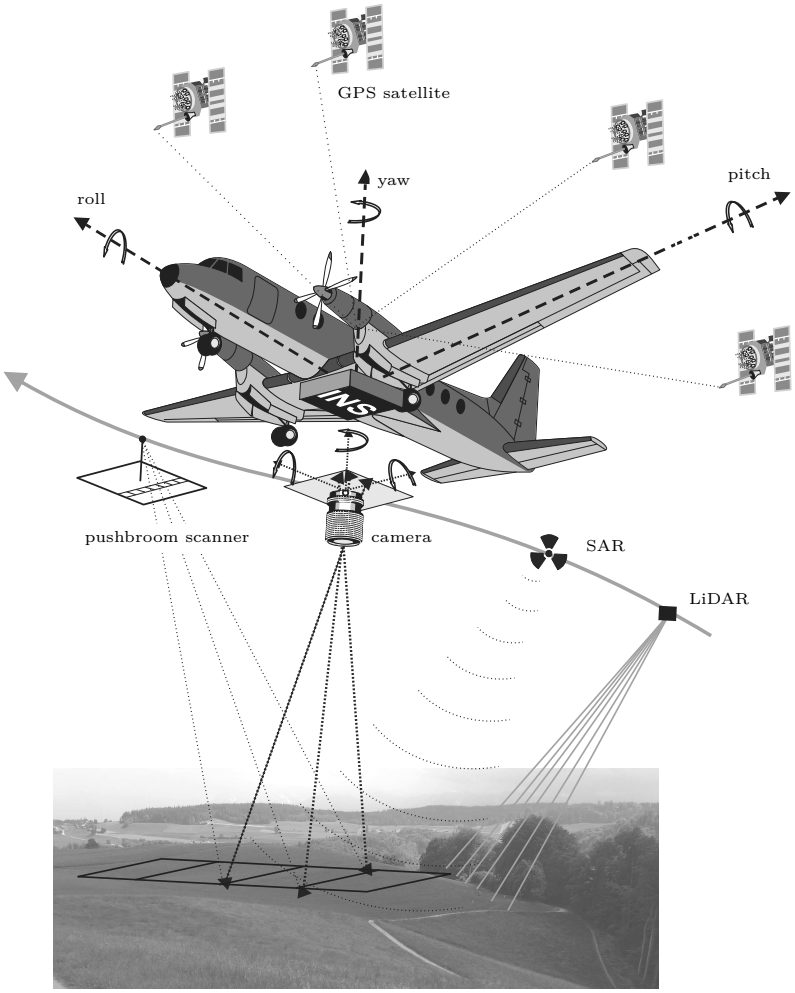


Figure 2.1: Use of navigation technology for sensor orientation.

Direct measurement of EO parameters typically relies on integrating receivers of the Global Navigation Satellite System (GNSS), such as GPS, with an inertial navigation system (INS) with the possible aiding of other sensors whose choice depends on the type of the carrier (e.g. odometers in cars or robots, barometers in aircrafts or drones, star-trackers on satellites). In a GNSS/INS system, GNSS data provides absolute position and velocity information as well as the error control of inertial measurements, while the INS contributes with attitude estimation, with the interpolation of the trajectory between GNSS position solutions and with the mitigation of sudden perturbations in GNSS measurements (e.g. cycle slips). Both technologies will be first introduced separately, while their integration will be described later (Sec. 2.4). The end of this section is devoted to the introduction of reference frame and to the establishment of relations for transferring the trajectory observation to sensors.

2.2 Satellite Navigation

2.2.1 Available systems

Satellite navigation have global or regional character (Fig. 2.2). There are four global navigation satellite systems (GNSS) put in place in chronological order by USA (GPS), Russia (GLONASS), Europe (Galileo) and China (Beidou-M). As of 2020 all systems are fully operational with somewhat similar constellations of 24 to 30 satellites (plus several spares) regularly organized into six (GPS) or three (others) orbital planes at medium Earth orbit (MEO). The slight differences in orbital altitude among constellations result in different orbital periods as denoted in Tab. 2.1. Through regular (GPS) or frequent (GLONASS and Beidou) satellite replacement and late deployment of Galileo, the open radio-navigation satellite service (RNSS) of each constellation uses either identical or very close frequencies and similar signal structures so the systems are interoperable.

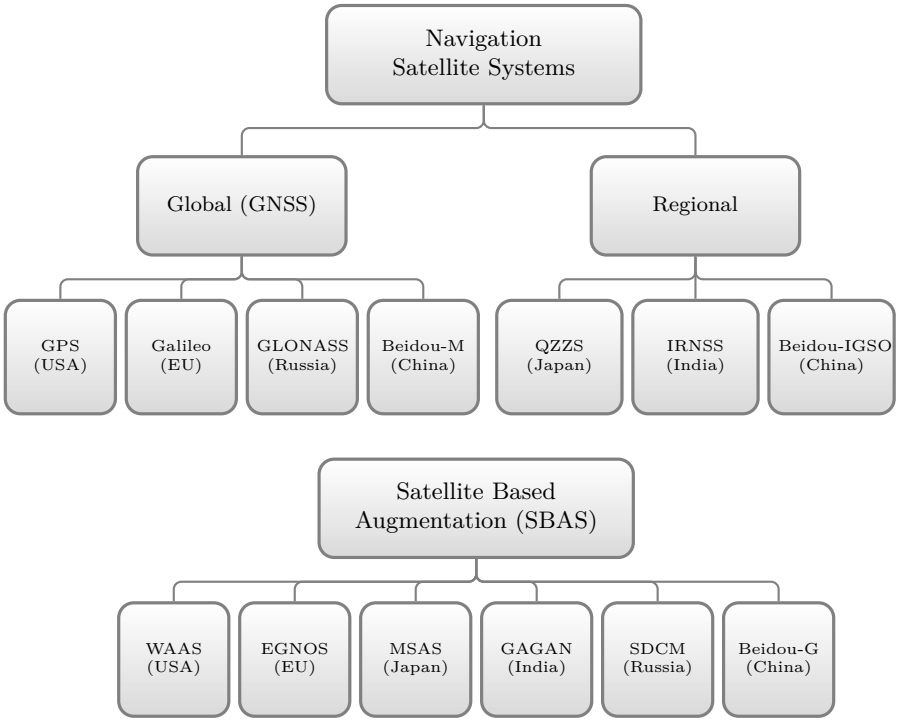


Figure 2.2: Top: Overview of today's GNSS. Left branch - satellites on global orbits, Right branch - satellites moving only above regions. Bottom: Satellite-based augmentation system with regional implementation of stationary satellites.

The regional satellite navigation employs Inclined Geosynchronous Orbit (IGSO) with 9 (QZZS), 7 (IRNSS) and 3 (Beidou-IGSO) satellites, re-

spectively. Apart emitting proprietary navigation services, these regional satellites broadcast also open RNSS to enhance GNSS availability over subcontinental areas. Similar enhancement is made by satellite-based augmentation system (SBAS) put forward by civilian aviation authorities. Using exclusively geostationary Earth orbits (GEO), SBAS is practically reaching a global coverage (bottom part of Fig. 2.2). Among them WAAS, EGNOS, MSAS, and GAGAN are certified as operational in meeting the exigence of employing GNSS for civilian aircraft navigation in terms of accuracy, integrity, continuity and availability. Apart functioning as additional GNSS satellites, SBAS monitors GNSS and provides timely warnings if their signals do not meet the required specifications. Especially for receivers operating on a single-frequency SBAS significantly improves the accuracy of height determination over the monitored/certified regions.

GNSS	**GPS**	**Galileo**	**GLONASS**	**Beidou-M**
Number of satellites	24-36	24-30	24-30	24-30
Orbital planes	6	3	3	3
Orbital altitude (km)	20,200	23,222	19,100	21,400
Orbital period (h:m:s)	11:58:02	14:04:41	11:14:30	12:52:04

Table 2.1: GNSS nominal constellation characteristics, after Betz (2016).

2.2.2 Signals structures

The situation of signals on current satellite navigation systems is complex due to evolution. As for GPS, the early satellites broadcasted signals only on two frequencies L1 (centred at 1575.42 MHz) and L2 (centred at 1227.60 MHz), while the modernized GPS includes also L5 (centered at 1176.45 MHz). The signal that remain open to all users on all generation of GPS satellites is the coarse-acquisition (C/A) code transmitted on L1. The later generation of GPS satellites emit additional open signal (L2C) on L2 frequency, while the modernized GPS added open L5 signal on a third frequency and open L1C signal on the first frequency (Table 2). The signal complexity increases from C/A over L2C to L5 and L1C with the goal of improving ranging accuracy, increasing robustness, mitigating adverse effects as multipath while improving interoperability with Galileo and other system.

To distinguish signals coming from different satellites GPS, Galileo and Beidou adopted a code division multiple access (CDMA), while GLONASS used frequency separation (FDMA) that made the fabrication of receiver more complex. To improve the interoperability with other systems, modernized GLONASS added CDMA on three frequencies while keeping FDMA for continuity. As shown in Tab. 2.2, the open service with CDMA on GLONASS is, however, available only on two frequencies. The situation is

somewhat similar for Galileo and Beidou that both adopted complex message structures on E5 and B2 that result in low noise level in code-based ranging (< 0.1 m). The full benefit of all these signals comes to its full potential when broadcasted by a large part of the satellites in every constellation. Thanks to the interoperability among GNSS the number of available satellites increased substantially over last the last decade. In addition, a combined single frequency GPS/Galileo/GLONASS/Beidou receiver is not significantly more expensive to manufacture than for one system. As explained further, receivers accessing signals on additional frequencies improve further the accuracy and reliability of satellite-based positioning.

Frequency (MHz)	**1176-1207**	**1227**	**1560-1600**
GPS	L5	L2C	C/A, L1C
Galileo	F5 (a+b)		E1
GLONASS	L3OC		L1OC
Beidou	B2 (a+b)		B1-C

Table 2.2: Open signals of modern GNSS based on CDMA.

2.2.3 Positioning methods

An overview of current GNSS positioning techniques is provided in Fig. 2.3. The selection of a particular method depends on the factors of accuracy and rapidity in data acquisition and mobile mapping. These methods are:

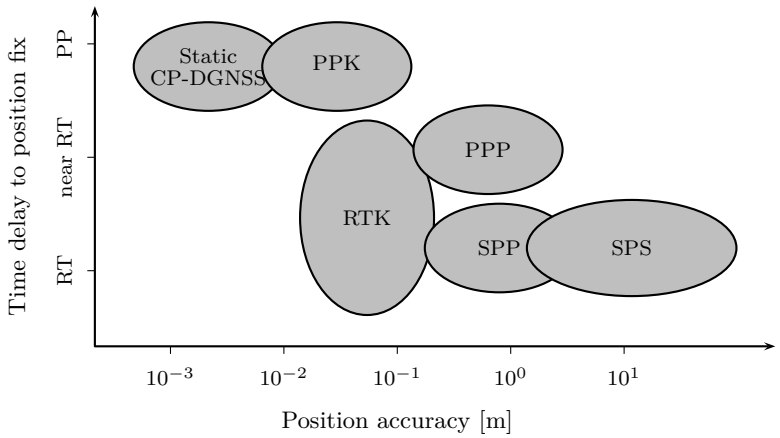


Figure 2.3: Overview of GNSS positioning methods as a function of accuracy and rapidity.

- *Single Point Positioning (SPP)* is the most commonly used method for real-time positioning. It is based on a single receiver and phase-smoothed code data processing (absolute GNSS positioning, (2) in

Fig. 2.3). Provided that SBAS corrections for altitude are available, this approach can deliver accuracies of 0.5 - 3 m. However, SBAS is not available world-wide and the reception of the ionospheric correction grid emitted by the geostationary satellite (that orbits above the equator) is valid only inside the monitored region. In such case the *Standard Positioning Service (SPS)* ((1) in Fig. 2.3) provides an accuracy of about 2 to 10 m.

- *Precise Point Positioning (PPP)* is a novel positioning methodology based on the fast availability (i.e., within an hour) of precise GNSS satellite orbit parameters and clock corrections, ((3) in Fig. 2.3). This technique can achieve sub-decimetric position accuracy (Satirapod and Homniam, 2006) and is available world-wide without the need of an augmentation system.

- *Differential* GNSS (DGNSS), *carrier-phase* DGNSS and *post-processed kinematic* (PPK) are relative positioning techniques based on simultaneous observations by the rover and base (one or or more) receivers, where the latter is placed at a location with known coordinates. DGNSS uses only the code (or carrier-smoothed code) observations, while the other two employ also the more precise but ambiguous carrier-phase measurements. The ambiguities are resolved via complex processing whose reliability is increased with dual-frequency observations. For *static* carrier-phase DGNSS ((6) in Fig. 2.3), sub-centimetre to millimetre accuracy can be achieved when respecting some considerations about baseline length and observation time. The upper limit in PPK ((5) in Fig. 2.3) is centimetre to decimetre accuracy for relative baseline length limit of around 15 km when using carrier-phase observations on two (or more) frequencies. The accuracy of PPK without ambiguity resolution is generally not better than a few decimetres while ambiguity resolution using observations on a single frequency is limited to 1-2 km long baseline.

- *Real-time kinematics (RTK)* applies the above mentioned PPK principles in the real-time ((4) in Fig. 2.3). Its prerequisite is the establishment of a communication link transmitting reference measurements or correction parameters. Similarly to PPK, this information is provided from a base receiver or from a network of receivers. National-wide networks broadcasting such type of corrections are available in many regions and can be accessed via modern communication technologies (i.e., Internet and mobile telephone networks). This makes them employable even for kinematic data acquisition. Sub-decimetre to centimetre-level positioning accuracy can be achieved by this means in ideal conditions.

2.3 Inertial navigation

Inertial navigation derives position, velocity and attitude from the initial knowledge of these quantities and from the integration of the observed accelerations (more precisely, *specific forces*) and *angular velocities* along their motion. These observations are normally obtained from a minimum of three accelerometers and three gyroscopes that are orthogonally mounted within an inertial measurement unit (IMU). An IMU coupled with a navigation computer creates an inertial navigation system (INS). A detailed overview of gyroscope and accelerometer technology can be found in (Titterton and Weston, 1997; Jekeli, 2001), the following list is limited to the types common in direct sensor orientation.

2.3.1 Gyroscope technology

The gyroscopes usually represent the most expensive part of an IMU. Their accuracy affects significantly the overall navigation performance of an INS. Several types of gyros are used in sensor orientation:

- *Mechanical gyros*: These gyroscopes use the principle of conservation of the angular momentum: A mass is spun at high speed around its axis, and the reaction to external forces (called precession forces) acting on its spin due to casing rotation is measured. The most common rotational gyro employed for sensor orientation is the *dynamically tuned gyro (DTG)*. It is relatively small, affordable and provides excellent short-term accuracy.

- *Optical gyros*: Such gyros use the Sagnac effect that rises due to the fact that speed of light is conserved in rotating systems (Andersson et al, 1994). The most common types are *ring laser gyros (RLG)* and *fiber optical gyros (FOG)*. Both are used for the most accuracy-demanding applications. FOGs of lower category are employed in the wider context of sensor orientation.

- *Vibratory gyros*: These gyroscopes exploit the principle that an oscillating body preserves the plane of vibration in inertial space despite rotations. These sensors are usually less precise, however, they are smaller and cheaper to fabricate. They are often employed in airborne applications with middle to low accuracy requirements.

- *MEMS-gyros*: These tiny gyroscopes exploit different physical principles and come in varying sizes and quality through microelectromechanical system (MEMS) technology that produces small and inexpensive sensors. They are used in mass market, auto-motive, robotic and entry level navigation applications. They are indispensable on drones

for control and guidance. As their high end approaches in some aspects the quality of low-end FOGs they are useful for robotics and UAV-based acquisition.

2.3.2 Accelerometer technology

For accelerometers three relevant types have to be mentioned.

- *Force rebalance accelerometers* measure the electrical current that is proportional to the force needed to maintain a suspended proof mass at rest under acceleration. These are used in the most demanding autonomous or airborne applications (e.g., precise underground or indoor mapping of large structures or high-altitude flights).
- *Vibrating accelerometers* exploit the resonant frequency of a mass hanging on a vibrating string. The frequency of vibration varies when an acceleration acts in the direction of the string. Such accelerometers are often fabricated as high-end MEMS sensors.
- *MEMS-gyros* based on different physical principles. Their high-end type is employed in robotics and UAV-based acquisition.

2.3.3 Strapdown INS

In earlier INS realizations the inertial sensors were mounted on stabilized (gimbaled) platforms, thus mechanically isolated from the rotational motion of the carrier. The advances in digital processing made it possible to avoid gimbaled mounts. Nowadays, the inertial sensors are rigidly mounted (strapped-down) to its casing, hereby decreasing the complexity and cost of the system while increasing the dynamic range of motion that can be tracked. As the number of moving parts is reduced these systems are also smaller and more reliable. A strapdown INS is often fitted in the same casing together with an optical instrument and its orientation output can be used for sensor-head stabilization (e.g. Sec. 4.6).

2.4 Integrated navigation

2.4.1 Principle

Integrated navigation is a technique that combines data from several navigation systems or sensors with the aim to improve the accuracy and robustness of the estimated trajectory. In this respect, the satellite and inertial navigation have a very different but complementary behavior. The performance of a standalone INS is characterized by a time-dependent drift in the accuracy of the position, velocity and attitude estimates it provides. The rate

at which the navigation errors grow in time is governed predominantly by the accuracy of the initial alignment, noise and imperfections in the inertial sensors and its assembly, as well as the dynamics of the trajectory. Whilst improved positioning accuracy can be achieved through the use of more accurate sensors, this cannot match the GNSS-type precision in the long run. On the other hand, the GNSS positioning is conditioned by the requirement for line of sight to a number of satellites (four or more), which is difficult to maintain in all situations, especially in terrestrial mobile mapping or indoors. Therefore, the combination of both systems enhances the trajectory determination across the spectrum of motion.

Contrary to GNSS, the inertial navigation provides continuous data output for all trajectory parameters (i.e. position, velocity and attitude). Therefore, the integrated navigation principally stabilizes and refines INS output by estimating and correcting the systematic effects in the inertial sensors and in the initialization. Different types of navigation aiding may be categorized as follows.

- *External measurements*: Measurements obtained by receiving signals or by viewing objects outside the vehicle. Such observations may be provided by radio navigation aids, GNSS satellites, star trackers or imagery, for example.
- *Autonomous measurements*: Measurements derived using additional sensors carried on-board without the dependence of external infrastructure or visibility. Navigation of this type may be provided by odometers, pressure sensors, Doppler radar or magnetic sensors, for example.
- *Dynamic constraints*: Application of implicit knowledge of some dynamical state or its form. For example, constraints such as zero velocity and non-holonomic condition (i.e. the alignment of vehicle speed with its direction) are used as supplementary aiding method on terrestrial vehicles or complete vehicle dynamic model is used for UAVs (Khaghani and Skaloud, 2016, 2018).

2.4.2 Integration schemes

Optimal integration of different measurement data with inertial observations is commonly achieved by using a Kalman filter/smoother (Fig. 2.4). The data filtering/smoothing can be, however, organized in different manners with respect to GNSS observations. The following two integration schemes are the most important[1]:

[1] A performance comparison between the two presented integration schemes can be found in Weiss and Kee (1995); Wei and Schwarz (1990) with the focus on RTK in Scherzinger (2006)

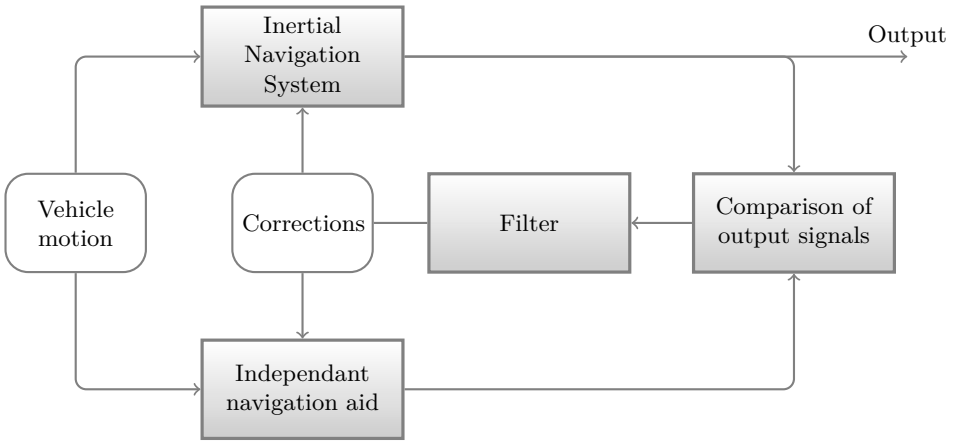


Figure 2.4: Typical integrated navigation scheme for direct georeferencing.

- *Loosely-coupled integration*: This is the most common integration approach, especially in airborne or shipborne installations. The raw IMU measurements are integrated to yield position and attitude at the IMU output rate (normally 100 to 500 Hz). The position and velocity data gathered by GNSS are processed independently, yielding a sequence of positions and velocities at a certain frequency (normally 0.1 to 1 Hz). These data are subsequently fed as updates within an extended Kalman Filter (EKF). The observed differences between the predicted (INS) and GNSS-determined velocities and positions are used to estimate the elements of the filter state vector, containing on one side the error states related to the trajectory (i.e. position, velocity and orientation) and on the other side those related to the inertial sensors themselves (i.e. gyro and accelerometers biases and scale factors, odometer or pressure sensor bias, etc.).

- *Closely-coupled integration*: In this integration scheme the GNSS raw measurements (normally double-differenced code, phase and Doppler measurements) are fed directly into the Kalman filter. Therefore, the GNSS measurements can be used in the filter even if the number of visible satellites is not sufficient to compute an independent position fix (i.e. lower than four). Accordingly, this integration scheme is advantageous for environments with reduced GNSS signal receptions (e.g. urban canyons), and is commonly used in terrestrial mobile mapping.

2.4.3 Resulting accuracy

For a strapdown INS, sensor integration solves firstly the problem of calibrating the systematic errors (i.e. residual gyro and accelerometer biases, scale factors etc.). Secondly, use of GNSS data mitigates attitude initial-

ization errors and in certain cases enables kinematic (in-flight alignment), which removes the need for the vehicle to be held stationary for the north-seeking process prior to movement. [2] At the same time, the inertial system smoothens the noisy velocity outputs from GNSS and provides high-rate measurement of position and velocity over larger spectrum of motion.

There is no such thing as a perfect instrument and as strong as it is, the integration cannot completely eliminate all errors. In other words, the data integration handled by a Kalman filter/smoother cancels only the non-overlapping part of the sensor's error budget. The performance of error cancelation depends on the motion of interest, the instrument type and the the encountered dynamics. While the long-term positioning accuracy limit depends on the GNSS positioning solution (i.e. Fig. 2.3), the time over which such accuracy can be maintained in the absence of satellite signals depends mainly on the quality of the INS and its preceding calibration. Based on the position error accumulated after 1 hour of autonomous operation, the INS are normally grouped into four main categories (Greenspan, 1995): *strategical-grade*, *navigation-grade*, *tactical-grade* and *low-cost* (MEMS) instruments. A summary of potential orientation accuracies for today's most popular sensors used for civilian applications in sensor orientation is summarized in Tab. 2.3. The automotive in Tab. 2.4 corresponds to small MEMS IMUs of high quality as those used by terrestrial and indoor robots or UAVs.

Time	**Navigation-grade**		**Tactical-grade**	
	roll/pitch (deg)	yaw (deg)	roll/pitch (deg)	yaw (deg)
1 sec	0.001-0.0014	0.001-0.002	0.002-0.02	0.001-0.05
1-3 min	0.0014-0.003	0.004-0.005	0.005-0.04	0.008-0.1
longer time	trajectory dependent - similar to 1-3 min when optimal			

Table 2.3: Orientation accuracy of as a function of time and INS quality.

Time	**Low-end tactical-grade**	
	roll/pitch (deg)	yaw (deg)
1 sec	0.005-0.1	0.005-0.2
1-3 min	0.03-0.2	0.1-0.2
longer time	trajectory dependent	

Table 2.4: Orientation accuracy for small MEMS IMU of high quality.

[2] This concerns all gyroscopes of lower accuracy as those employed in UAVs that cannot complete north-seeking without an external assistance.

2.5 Geometrical relations

2.5.1 Coordinate frames

Sensor frame - *s*

While the nature of imagery captured in a static environment (e.g. terrestrial laser scanning or terrestrial photogrammetry) allows operations within a local coordinate system, kinematic remote sensing with navigation support requires the employment of global reference frame as well as several intermediate frames. Tab. 2.5 provides an overview of the frames used, together with their abbreviated identifiers. All frames are defined to be right-handed Cartesian frames, and their relation is schematically depicted in Fig. 2.5 for a case of an airborne laser scanner. The situation is similar for other optical devices for which the sensor frames(s) are defined later (c.f., Sec. 1, Sec. 4).

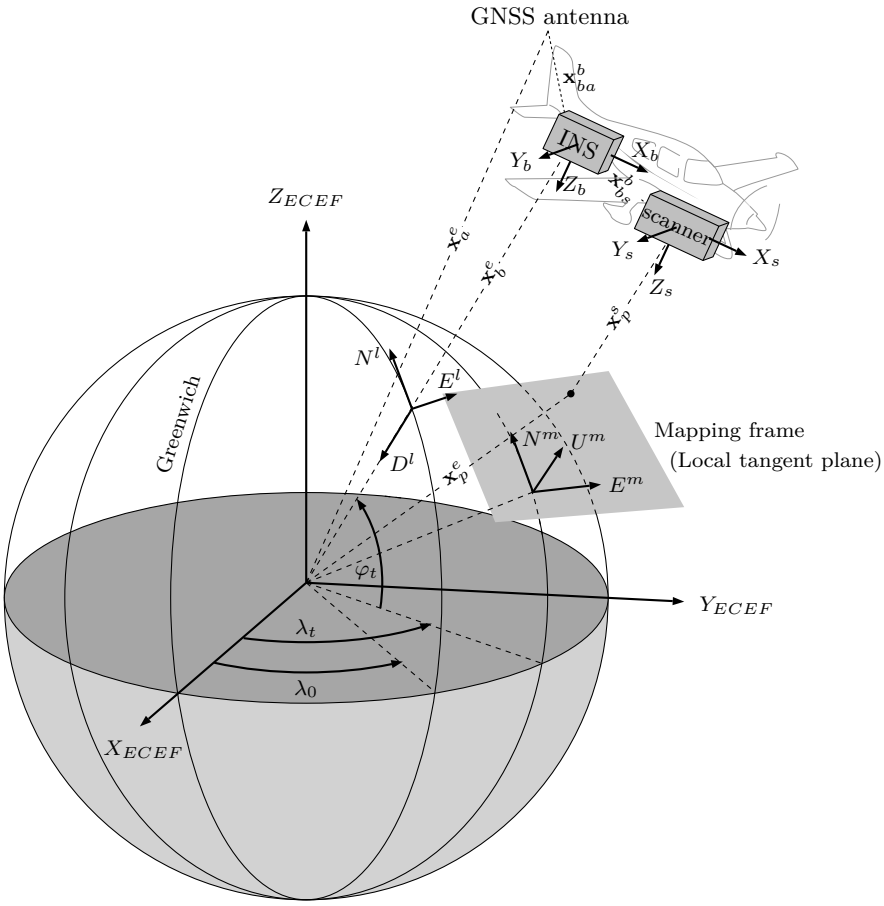


Figure 2.5: Geometry of direct sensor orientation.

ID	Frame name	Description
s	Sensor frame	Frame of the laser sensor, defined by the principal axes of an optical instrument; e.g. xy-axes define an image plane in the frame imagery, yz-defines the scanning plane of a 2D scanner
b	Body frame	Frame realized by the triad of accelerometers within an IMU
l	Local level frame	This frame is tangent to the global ellipsoid (normally WGS84), with the orthogonal components usually defined as N-orth (x), E-ast (y) and D -own (z)
e	ECEF frame	Earth-centered Earth-fixed frame. The origin is the geocenter of the Earth, x-axis points towards the Greenwich Meridian and the z-axis is the mean direction of the Earth rotation axis. The y-axis completes the right-handed Cartesian system
m	Mapping frame	Cartesian frame with E-ast (x), N-orth (y) and U-p (z) component. The easiest implementation is the local tangent plane frame, but this frame can also be represented by a projection and/or national datum

Table 2.5: Overview of reference frames.

Earth-Centered Earth-Fixed frame - *e*

The satellite orbits of the common GNSS-systems are referred to this frame and so is the outcome of the trajectory computation. A geocentric ellipsoid is normally attached to ECEF frame, and its properties together with other geophysical parameters define a world datum (e.g., WGS84 used for GPS measurements). Coordinates in this frame can either be expressed as *geocentric coordinates* (x^e ,y^e ,z^e), or as *geographical coordinates* (latitude φ, longitude λ, ellipsoidal height h). The latter parametrization is often used in the output of GNSS/INS trajectory. The relation between the Cartesian and ellipsoidal coordinates reads:

$$\mathbf{x}^e = \begin{pmatrix} x_1^e \\ x_2^e \\ x_3^e \end{pmatrix} = \begin{pmatrix} (N+h)\cos\varphi\,\cos\lambda \\ (N+h)\cos\varphi\,\sin\lambda \\ \left(\frac{b^2}{a^2}N+h\right)\sin\varphi \end{pmatrix}, \tag{2.1}$$

where N is the radius of curvature in the prime vertical and a and b are the semi-major and semi-minor axes, respectively.

Local-level frame - *l*

This frame is mainly used as the reference for the orientation angles output from the GNSS/INS processing. Its origin is defined by the sensor position

on a reference ellipsoid at zero height, which corresponds to the intersection of the local vertical at the actual sensor position with the reference surface. The $x^{\,l}-$axis points along the local meridian to the north , the $y^{\,l}-$axis points to the east and the $z^{\,l}-$ axis points down to complete the system. Such local-frame definition is called l-NED (for north−east−down), while the upward positive convention of the $z^{\,l}-$ axis defines the l-ENU frame (east−north−up). The rotation from the l- to the e-frame can be described by the matrix $\mathbf{R}^{e}_{l_{NED}}$:

$$\mathbf{R}^{e}_{l_{NED}} = \begin{pmatrix} -\sin\varphi\,\cos\lambda & -\sin\lambda & -\cos\varphi\,\cos\lambda \\ -\sin\varphi\,\sin\lambda & \cos\lambda & -\cos\varphi\,\sin\lambda \\ \cos\varphi & 0 & -\sin\varphi \end{pmatrix} . \tag{2.2}$$

Body frame - *b*

The body-frame is represented by the axes of the inertial navigation system. The origin of the b-frame is located at the navigation center of the INS and the axes are congruent with the axes spanned by the triad of accelerometers. Normally, the b-frame axis coincides with the principal axis of rotation of the carrier, or can be rotated to them by some cardinal rotation. According to the aerospace norm ARINC 705, the axis and the rotations describing the 3-D attitude are defined as follows. The $x^{\,b}-$axis is pointing forward along the fuselage, the $y^{\,b}$-axis points to the right, and the $z^{\,b}-$axis points down. The associated rotation angles along the $x-y-z$ axes are referred to as *roll* (r), *pitch* (p) and *yaw* (y). Respecting the aerospace attitude definitions, the corresponding rotation matrix that relates the l-frame to the b-frame takes the following form:

$$\mathbf{R}^{b}_{l_{NED}} = \mathbf{R}_x(r)\,\mathbf{R}_y(p)\,\mathbf{R}_z(y)\,. \tag{2.3}$$

where $\mathbf{R}_x(r)$, $\mathbf{R}_y(p)$ and $\mathbf{R}_z(y)$ are defined as:

$$\mathbf{R}_x(r) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos r & \sin r \\ 0 & -\sin r & \cos r \end{bmatrix}$$

$$\mathbf{R}_y(p) = \begin{bmatrix} \cos p & 0 & -\sin p \\ 0 & 1 & 0 \\ \sin p & 0 & \cos p \end{bmatrix} \tag{2.4}$$

$$\mathbf{R}_z(y) = \begin{bmatrix} \cos y & \sin y & 0 \\ -\sin y & \cos y & 0 \\ 0 & 0 & 1 \end{bmatrix} .$$

2.5.2 Transformation of exterior orientation

The relationship between an arbitrary point $\mathbf{x}_p^s$ in the s-frame coordinates and that same vector expressed in the b-frame is given by

$$\mathbf{x}_p^b = \mathbf{x}_{bs}^b + \mathbf{R}_s^b(\omega_b, \varphi_b, \kappa_b) \cdot \mathbf{x}_p^s \,, \tag{2.5}$$

where $\mathbf{x}_{bs}^b = \mathbf{R}_s^b \mathbf{x}_{bs}^s$ denotes the origin of the s-frame in the b-frame, which is also known as the lever-arm vector. The rotation matrix $\mathbf{R}_s^b$ in (2.5) is called the boresight and represents the relative misalignment between the s- and b-frames. This matrix is usually parameterized by the three Euler angles ω_b, φ_b, κ_b. The magnitude of the boresight angles and the lever-arm need to be determined by calibration.

The observation equation for direct sensor orientation for a point p viewed by a sensor s in the e-frame coordinates follows from Fig. 2.5 by combining (2.1) to (2.3):

$$\mathbf{x}_p^e(t) = \mathbf{x}_b^e(t) + \mathbf{R}_l^e(t)\mathbf{R}_b^l(t)\mathbf{R}_s^b(\omega_b, \phi_b, \kappa_b) \left(\mathbf{x}_{bs}^s + \mathbf{x}_p^s(t)\right) \tag{2.6}$$

where $\mathbf{x}_b^e(t)$ is the navigation center of the IMU in the e-frame and all other components were defined previously. The symbol (t) indicates quantities that vary with time.

Mapping frame - *m*

For active sensors as laser scanners, the coordinates of observed points in ECEF frame can be generated via (2.6). However, the final coordinates are often needed in some other datum and projection. The so-called mapping frame habitually represents a national coordinate system, and the results of mapping can be transferred to such a frame point- or pixel-wise via relations published by local surveying authorities. Alternatively, the registration of optical images and that of laser can be performed directly in mapping frame as discussed in detail in Legat (2006); Skaloud and Legat (2008).

2.5.3 System calibration

The method of direct sensor orientation requires that the optical sensor be calibrated for the parameters of interior orientation, which includes, system installation. The latter concerns determining the spatial and orientation offsets that exists between optical and navigation sensors.

The lever-arm $\mathbf{x}_{bs}^b$ is either specified by the system provider or needs to be determined per installation. The same is true for the lever-arm between the IMU center and the GNSS antenna $\mathbf{x}_{ba}^b$, which is needed during GNSS/INS integration. Calibration of the lever-arms is best performed by tacheometry. Such procedure is discussed in detail by Schaer (2009); Rehak

and Skaloud (2015) for an aircraft and small UAV system, respectively. An alternative solution is to estimate $\mathbf{x}_{ba}^{b}$ directly within the GNSS/INS Kalman filter/smoother as an additional parameter. Similarly, $\mathbf{x}_{bs}^{b}$ can be estimated in the block adjustment (Sec. 4), but its value is often strongly correlated to other parameters and this approach should be therefore avoided when later used for direct sensor orientation.

The recovery of the the boresight matrix $\mathbf{R}_{s}^{b}$ is more involved and requires the use of principles described in Sec. 4. For frame cameras this process can be achieved either in one (Cramer and Stallmann, 2002; Kruck, 2001) or two steps (Skaloud and Schaer, 2003). A similar procedure is maintained for the line scanners (Cosandier, 1999; Tempelmann and Hinsken, 2005). The boresight determination in kinematic laser scanning followed a rapid evolution (Burman, 2000; Kager, 2004; Morin and El-Sheimy, 2002) that converged to the approach based on surfaces of known form (Friess, 2006; Skaloud and Lichti, 2006). The calibration principles are further addressed in Sec. 4.

Chapter 3
Photogrammetry

3.1 From 2D to 3D

The main task of photogrammetry or equivalently *computer vision* is to reconstruct 3-D scene from 2-D images. The most important requirement for the reconstruction to work is the that the scene is imaged from different places so that sufficient correspondences between pictures can be (automatically) established. Position and orientation of each image is found along the way, fact of which allow to infer the motion of the camera and thus the platform (up to the image acquisition rate). At the same time, the knowledge of camera motion observed by other sensor(s) as those discussed in Sec. 2 can be used in support of the reconstruction process, which is a subject of Sec. 4.

Given set of images, the principals challenges (and steps) of reconstructing 3D models are threefold (Fig. 3.1)

- Correspondences: automatically detect sufficient number of *key* features on each image and establish their correspondences with other images (Sec. 3.8)
- Geometry (motion, orientation): recover camera pose (position and orientation) between images, its intrinsic parameters (calibration) and feature's 3D coordinates
- Scene (structure): using the knowledge of geometry, create dense point cloud to recover 3D objects (models) with texture (Sec. 3.8.3)

This section essentially concentrate on the geometry part of the reconstruction problem without the help of navigation sensors. It is known under different names as *orientation and calibration* in photogrammetry or *structure from motion* in computer vision as well as *bundle adjustment* (both).

We will describe the process of image formation using a mathematical model that accounts for three types of transformations:

1. coordinate transformation between image coordinate frames;
2. projection of a 3-D scene onto 2-D image coordinates;
3. relation between the camera frame and an external mapping frame.

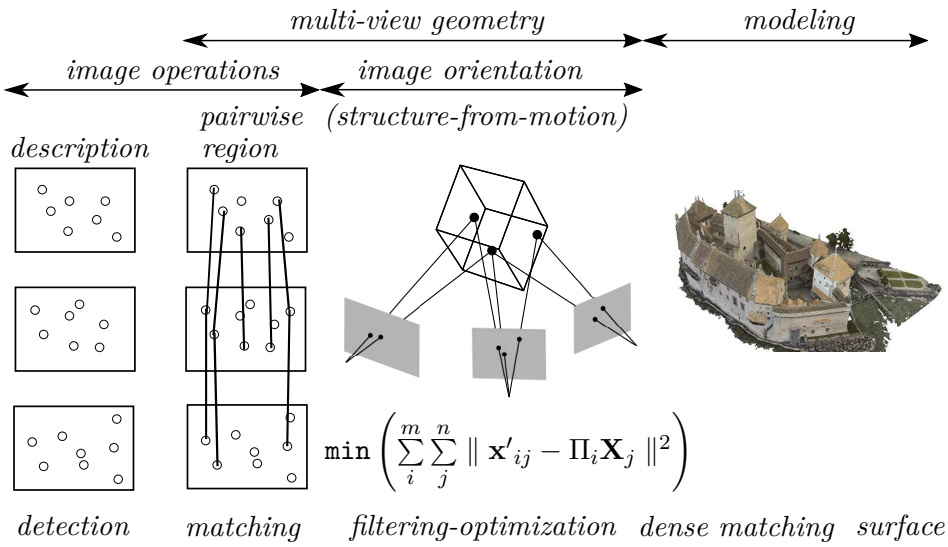


Figure 3.1: Photogrammetry/computer-vision process for 3-D scene reconstructions based on 2-D imagery.

This chapter proceeds as follows. Although a general introduction to the relation between sensor-mapping frame was given in Sec. 2.5, we repeat it here but in homogeneous representation of coordinates (Sec. 3.2), form of which will be useful later on. Then, after introducing a basic geometry of the imaging system (Sec. 3.3 - 3.4) we describe a model of image formation for an ideal perspective camera (Sec. 3.5. With the necessary components we introduce the reconstruction process for a stereo-pair of images (Sec. 3.6) that we later extend for multiple views (Sec. 3.7). We conclude the chapter with different processing strategies for filtering and optimization in scene reconstruction.

3.2 Camera pose in a homogeneous form

Consider two Cartesian frames, where one is a mapping frame spanning the object space, and the second one is related to a camera, viewing a scene at certain time t, to which belongs a point p. From the situation depicted in Fig. 3.2 it is clear, that the coordinates of a point p with respect to the mapping frame m is simply the sum of the translation $\mathbf{x}_c^m$ of the origin of the frame c relative to that of the frame m and the vector $\mathbf{x}^c$ expressed in relation to the mapping frame m, which is $\mathbf{R}_c^m \mathbf{x}^c$, where $\mathbf{R}_c^m$ is the relative rotation between the frames

$$\mathbf{x}^m = \mathbf{x}_c^m + \mathbf{R}_c^m \mathbf{x}^c \,. \tag{3.1}$$

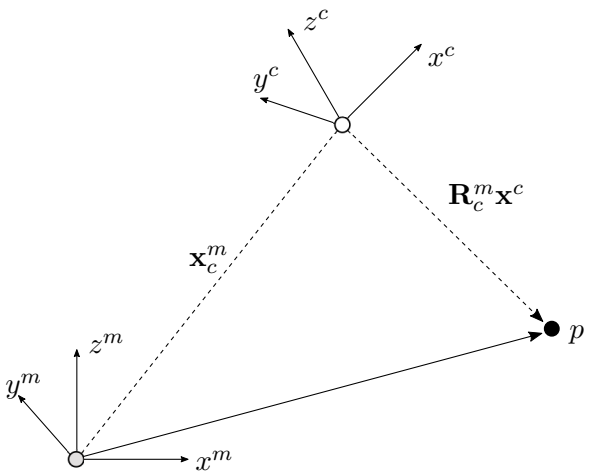


Figure 3.2: Motion of camera frame with respect to a Cartesian mapping frame.

Every time the camera moves, its motion is captured by $\mathbf{T}_c^m = (\mathbf{R}_c^m, \mathbf{x}_c^m)$ or more shortly by $\mathbf{T} = (\mathbf{R}, \mathbf{x})$ when the involved frames are clear from the context. It will become an advantage when we convert the transformation expressed by (3.1) to an expression of a form $\mathbf{u} = \mathbf{A}\mathbf{v}$. This is possible by adding "1" to the vector x as its fourth coordinate and by defining operations on so called *homogeneous coordinates*. Such extension preserves the original Euclidean space.

$$\overline{\mathbf{x}} \doteq \begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix} = \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{3.2}$$

The vectors are defined analogically as differences of coordinates $\overline{v} = \overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2$. Differences makes the fourth component null and give rise to the original subspace. Rewriting (3.1) in the new notation leads to

$$\overline{\mathbf{x}}^m = \begin{pmatrix} \mathbf{x}^m \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{R}_c^m & \mathbf{x}_c^m \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{x}^c \\ 1 \end{pmatrix} \doteq \overline{\mathbf{T}}_c^m \, \overline{\mathbf{x}}_c \tag{3.3}$$

where the 4×4 matrix $\overline{\mathbf{T}}_c^m$ is the *homogeneous representation* of the rigid-body transformation. Now is possible to encapsulate the coordinate transformation between several frames as a sequence of multiplications

$$\overline{\mathbf{T}}_c^a = \overline{\mathbf{T}}_b^a \overline{\mathbf{T}}_c^b = \begin{pmatrix} \mathbf{R}_b^a & \mathbf{x}_b^a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{R}_c^b & \mathbf{x}_c^b \\ 0 & 1 \end{pmatrix}. \tag{3.4}$$

As can be easily verified, the inverse transformation is

$$\overline{\mathbf{T}}^{-1} = \begin{pmatrix} \mathbf{R} & \mathbf{x} \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{R}^T & -\mathbf{R}^T\mathbf{x} \\ 0 & 1 \end{pmatrix} \tag{3.5}$$

3.3 Pinhole camera

Consider the basic imaging system as described by Fig. 1.1 in Sec. 1.1. If the aperture of the lens decreases to zero, the only feature that contributes to illumination of an image point is that on the line going through the center of the lens o. This way an *image point* can be directly related to an *object point* as shown in the upper part of Fig. 3.3, where the camera frame is for the simplicity oriented in the same direction as the object frame and the point p is such that its image coordinate $y' = 0$.

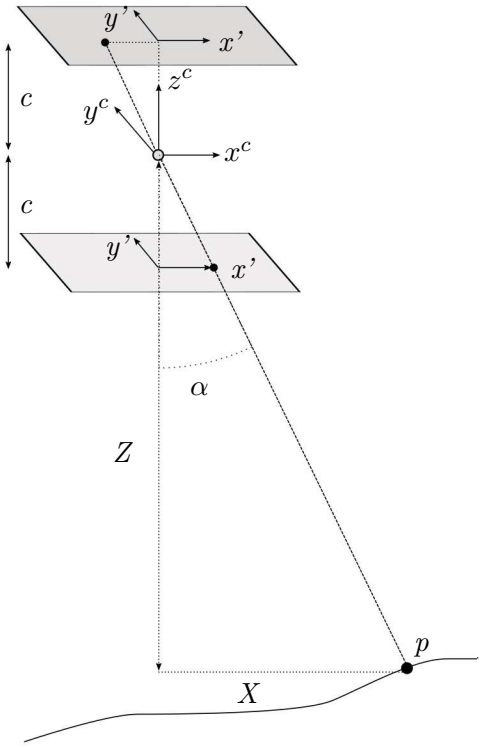


Figure 3.3: Pinhole imaging model and its frontal counterpart: the 3-D point p is projected on the image at the intersection of the ray going through the optical center o and the image plane at a distance c. Note that $\tan(\alpha) = X/Z$.

Lets define the distance from the object point to the optical center along the optical axis as Z, and the "horizontal' projection of the point on the

optical axis as X, while Y completes the right-handed system. On the image side, the distance from the optical center to the image plane along the optical axis is the camera constant c, while the distances from the intersection between the optical axis with the image plane to the image point are $-x'$ and $-y'$, respectively. From the similarity of the triangles in the upper part of Fig. 3.3 the coordinates of the image are related to that of the object by *perspective projection*.

$$x' = -c\,\frac{X}{Z}\,, \quad y' = -c\,\frac{Y}{Z} \tag{3.6}$$

The negative sign in (3.6) makes the object appear upside down on the image plane. Such reversing of the scene by perspective geometry of the lens is normally compensated by the optical system and we can eliminate this effect also mathematically by flipping the image coordinates $(-x',\, -y') \longmapsto (x',\, y')$. This is represented on the lower part of the Fig. 3.3 as virtually displacing the image plane in front of the optical center, which is so called *frontal pinhole camera model*. We define the camera frame with x' and y' axis identical to the "frontal" image plane, while placing its origin in the optical center. The z axis completes the right-handed coordinate system and its positive direction may go either "towards" or "away" from the object according to the arbitrary choice of image coordinates. Applying the change of the sign to (3.6) and combining both coordinates into a vector yields

$$\mathbf{x}' = \begin{pmatrix} x' \\ y' \end{pmatrix} = \frac{c}{Z} \begin{pmatrix} X \\ Y \end{pmatrix}, \tag{3.7}$$

or equivalently in a homogeneous form

$$Z \underbrace{\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix}}_{\bar{\mathbf{x}}'} = \begin{pmatrix} c & 0 & 0 & 0 \\ 0 & c & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \underbrace{\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}}_{\bar{\mathbf{x}}^c}, \tag{3.8}$$

where $\overline{\mathbf{x}}' \doteq (x', y', 1)^T$ and $\overline{\mathbf{x}}^c \doteq (X, Y, Z, 1)^T$ are homogeneous representation of image and camera coordinates, respectively. Note also that the unit of c is the same as of x', y'.

3.4 Image coordinates

Considering a digital camera, the measurements of features or "points" on the sensor are expressed in pixels. The usual convention is to situate the origin of pixel counting to the upper left corner of the image and express its coordinates in terms of rows and columns. However, we need to relate

the pixels to the frontal pinhole camera model. As depicted in Fig. 3.4 the optical axis intersects the sensor at *principal point* (PP). The principal point is usually close to the physical center of the sensor denoted as the *principal point of symmetry* (PPS). Based on these different origins we define three image coordinates systems, units of which are specified in Tab. 3.1.

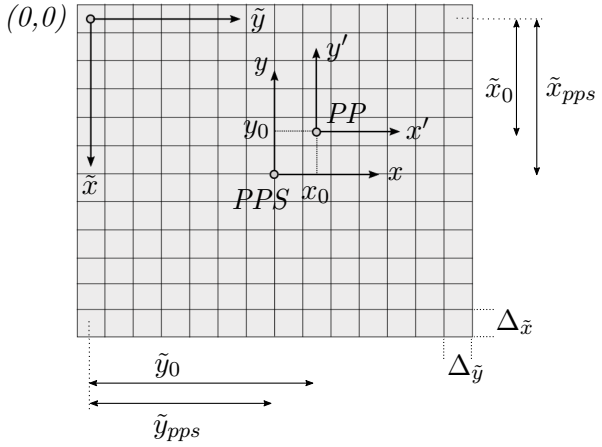


Figure 3.4: Pixel $(\tilde{x}, \tilde{y})$, sensor-centered (x, y) and perspective-centered (x', y') image coordinates.

	Origin	**Units**	**Usage**
$(\tilde{x}, \tilde{y})$	rows/cols counter	pixels	computer-vision (CV)
(x, y)	PPS	mm / pixels	photogrammetry / CV
(x', y')	PP	unitless (=1)	general

Table 3.1: Definition of different image coordinate systems.

The transformation from pixel rows and columns $(\tilde{x}, \tilde{y})$ to a metric, sensor-centered image coordinates (x, y) with axis orientation as in Fig. 3.4 considers the position of PPS in pixels $(\tilde{x}_{pps}, \tilde{y}_{pps})$ and pixel size (e.g., in mm) along rows and columns $(\Delta_{\tilde{x}}, \Delta_{\tilde{y}})$:

$$\begin{aligned} x &= \left[+\tilde{y} - \left(\tfrac{n_c}{2} - \tfrac{1}{2}\right)\right] \Delta_{\tilde{y}}\,, \\ y &= \left[-\tilde{x} + \left(\tfrac{n_r}{2} - \tfrac{1}{2}\right)\right] \Delta_{\tilde{x}}\,, \end{aligned} \tag{3.9}$$

where n_c and n_r is the total number of rows and columns, respectively. We express also the inverse relation, this time in a homogeneous form

$$\begin{pmatrix} \tilde{x} \\ \tilde{y} \\ 1 \end{pmatrix} = \begin{pmatrix} s_x & 0 & \tilde{x}_{pps} \\ 0 & s_y & \tilde{y}_{pps} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -y \\ x \\ 1 \end{pmatrix}. \tag{3.10}$$

with $\tilde{x}_{pps} = (n_r - 1)/2, \tilde{y}_{pps} = (n_c - 1)/2$ given in pixels and $s_x = 1/\Delta_{\tilde{x}}$, $s_y = 1/\Delta_{\tilde{y}}$.

To respect the perspective geometry we define a coordinate system (x', y') with an origin placed at the principal point of auto-collimation (PP). The orientation of the axis is arbitrary, but in photogrammetry is usually defined as in Fig. 3.4. We also chose the unit of this coordinate system to be equal to the principal distance c, so its coordinates correspond to the tangent of angles as shown in Fig. 3.3 for x' and α. The transformation from so called *reduced coordinates* (x', y') back to sensor centred coordinates (x, y) in a homogeneous form is

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} c & 0 & x_0 \\ 0 & c & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix}, \tag{3.11}$$

with x_0, y_0 expressed in mm (sometimes displayed in μm with 1/1000 scaling factor). Analogically, the transformation from (x', y') to rows/columns $(\tilde{x}, \tilde{y})$ in pixels is

$$\begin{pmatrix} \tilde{x} \\ \tilde{y} \\ 1 \end{pmatrix} = \begin{pmatrix} \tilde{c} & 0 & \tilde{x}_0 \\ 0 & \tilde{c} & \tilde{y}_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -y' \\ x' \\ 1 \end{pmatrix}, \tag{3.12}$$

with $\tilde{c}, \tilde{x}_0, \tilde{y}_0$ expressed in pixels.

Tab. 3.2 summarizes the transformations between the respective image coordinated systems.

	$(\tilde{x}, \tilde{y})$	(x, y)	(x', y')
$(\tilde{x}, \tilde{y})$		(3.10)	(3.12)
(x, y)	(3.9)		(3.11)

Table 3.2: Relations between image coordinate systems.

3.5 Imaging formation model

We now relate the mapping/object coordinates of point p with its coordinates on the image by means of perspective projection, while utilizing the camera frame along the way. Let us recall from Sec. 3.2 that the mapping coordinates of a point $\mathbf{x}^m = (X^m, Y^m, Z^m)^T$ relative that of a camera $\mathbf{x}^c$ are related by the rigid body transformation (inverse of 3.3)

$$\overline{\mathbf{x}}^c = \overline{\mathbf{T}}_c^{m-1} \overline{\mathbf{x}}^m, \tag{3.13}$$

where the homogeneous transformation $\overline{\mathbf{T}}$ contains both, the rotation and translation parameters $(\mathbf{R}, \mathbf{x})$.

Adopting the frontal camera model introduced in Sec. 3.3 for sensor-centered image coordinates (x, y) we rewrite (3.8) in vector notation

$$Z\,\overline{\mathbf{x}} = \begin{pmatrix} c & 0 & x_0 & 0 \\ 0 & c & y_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \overline{\mathbf{x}}^c \tag{3.14}$$

Since the depth of the point p represented by Z coordinate is unknown on a single photograph we may express it as one multiplied by an arbitrary scalar μ, i.e. $Z = \mu \cdot 1$. Decomposing the matrix in (3.14) into a product of two matrices while substituting for $\overline{\mathbf{x}}^c$ on the right side with (3.13) we obtain the geometric model for *a basic camera*

$$\mu \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} c & 0 & x_0 \\ 0 & c & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{R}_c^m & \mathbf{x}_c^m \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} X^m \\ Y^m \\ Z^m \\ 1 \end{pmatrix}. \tag{3.15}$$

By defining the first two matrices on the right-hand side of the above equation as

$$\mathbf{K} \doteq \begin{pmatrix} c & 0 & x_0 \\ 0 & c & y_0 \\ 0 & 0 & 1 \end{pmatrix}, \; \Pi_0 \doteq \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \tag{3.16}$$

we can rewrite the relation for a basic camera model (3.15) in a matrix form

$$\mu\overline{\mathbf{x}} = \mathbf{K}\Pi_0\overline{\mathbf{x}}^c = \mathbf{K}\Pi_0 \left(\overline{\mathbf{T}}_c^m\right)^{-1} \overline{\mathbf{x}}^m = \Pi\overline{\mathbf{x}}^m, \tag{3.17}$$

when combining the 3 x 4 matrix $\mathbf{K}\Pi_0\overline{\mathbf{T}}^{-1}$ into a general *projection matrix* Π.

Now we can consider also other intrinsic parameters of a camera, for example, a basic distortion of perspective-centred image coordinates (x'_d, y'_d) with radial symmetry as

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \left(1 + a_1\, r^2 + a_2\, r^4\right) \begin{pmatrix} x'_d \\ y'_d \end{pmatrix} \tag{3.18}$$

where $r^2 = x'^2_d + y'^2_d$ is the square of the distance from the principal point of autocollimation and a_1, a_2 are the *distortion coefficients*. When needed, this simple radial model can be extended by additional coefficients, as in (4.1). Combining the relation of simple image distortion (3.18) together with basic camera projection model (3.17) we define the *realistic image formation model* that is applicable to many cameras employed in photogrammetry.

3.6 Scene from two views

The previously described image formation model (3.17) $\mu\overline{\mathbf{x}}' = \Pi\overline{\mathbf{x}}^m$ relates the object coordinates to image coordinates. Now we would like to perform the inverse - reconstruct 3D object coordinates from images. As the scale (depth) μ is generally unknown due to $3D \longmapsto 2D$ projection (note that μ varies per point and image), we need to employ at least two images of the same object with different camera pose that are known. Such situation is depicted in Fig. 3.5: 3-D can be obtain by intersecting the couple of vectors pointing to the same object from two cameras. As suggested by the picture, the vector direction follows from image observation and internal camera geometry, however, both vectors need to refer to a common coordinate frame. This is the same as relating the respective camera poses to such a frame. Hence, the camera poses need to be found first. How this can be done using image observation only is described in the following.

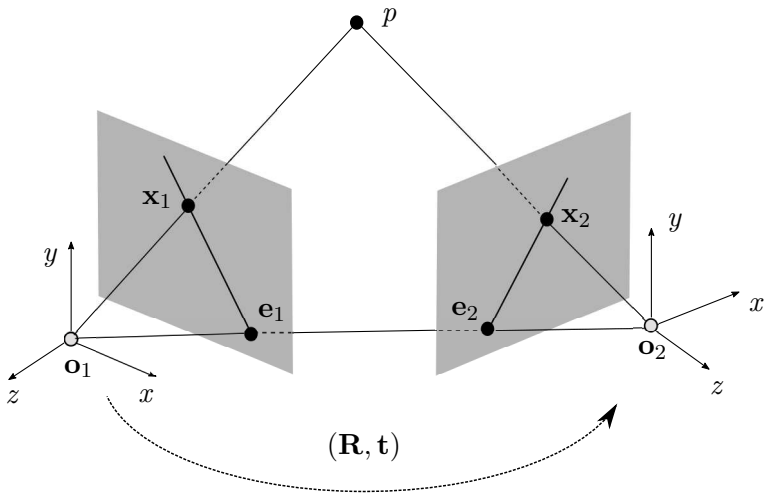


Figure 3.5: Main steps of photogrammetry/computer-vision process in 3-D scene reconstructions based on 2-D imagery.

3.6.1 Coplanarity constrain

We can relate the camera pose $\overline{\mathbf{T}}_2$ to the first one $\overline{\mathbf{T}}_1$ in a relative sense $\overline{\mathbf{T}}\,(\mathbf{R},\mathbf{t}) = \left(\overline{\mathbf{T}}\,{}^{m}_{c1}\right)^{-1}\left(\overline{\mathbf{T}}\,{}^{m}_{c2}\right)$, so that $\overline{\mathbf{x}}^m_{c,2} = \overline{\mathbf{T}}\,\overline{\mathbf{x}}^m_{c,1}$. Expressing this in image coordinates with respect to PP we obtain

$$\mu_2 \mathbf{x'}_2 = \mathbf{R}\mu_1 \mathbf{x'}_1 + \mathbf{t} \tag{3.19}$$

To eliminate the unknown depth μ we make a couple of steps. First, we multiply both sides of the above equation from left by a skew-symmetric[1] matrix $[\mathbf{t}\times]$ containing vector $\mathbf{t}$

$$\mu_2 [\mathbf{t}\times] \mathbf{x'}_2 = [\mathbf{t}\times] \mathbf{R} \mu_1 \mathbf{x'}_1 + [\mathbf{t}\times] \mathbf{t} \,. \tag{3.20}$$

Due to orthogonality, the last term on the right-hand side is zero. Second, we multiply the last relation by $\mathbf{x'}_2^T$

$$\mu_2 \mathbf{x'}_2^T [\mathbf{t}\times] \mathbf{x'}_2 = \mathbf{x'}_2^T [\mathbf{t}\times] \mathbf{R} \mu_1 \mathbf{x'}_1 \,. \tag{3.21}$$

Since $[\mathbf{t}\times] \mathbf{x'}_2$ is perpendicular to $\mathbf{x'}_2$ and the inner product of the two perpendicular vectors is zero, the left side $\mathbf{x'}_2^T [\mathbf{t}\times] \mathbf{x'}_2 = 0$. Also, as $\mu_1 \neq 0$ we can write

$$\mathbf{x'}_2^T [\mathbf{t}\times] \mathbf{R} \mathbf{x'}_1 = \mathbf{x'}_2^T \mathbf{E} \mathbf{x'}_1 = 0 \,, \tag{3.22}$$

where $\mathbf{E} \doteq [\mathbf{t}\times] \mathbf{R}$ is called the *essential matrix*. The above relation is called *epipolar constraint* as it conditions the three vectors $\mathbf{x'}_2$, $\mathbf{t}$ and $\mathbf{R}\,\mathbf{x'}_1$ to lie on a common plane, denoted as *epipolar plane*. Fig. 3.5 depicts also the two *epipols* $\mathbf{e}_1, \mathbf{e}_2$ resulting from the intersection between a line $\mathbf{o}_1 - \mathbf{o}_2$ and respective image planes. Connections $\mathbf{e}_1 - \mathbf{x'}_1$ and $\mathbf{e}_2 - \mathbf{x'}_2$ are called *epipolar lines* (intersections between the epipolar plane and the two image planes).

3.6.2 Essential matrix determination

To reconstruct $\mathbf{E}$ using only image observations, we briefly present the basic algorithm of Longuet-Higgins (1981) known also as the 8-point algorithm.

First, we stack the 3×3 entries of $\mathbf{E}$ into a vector by columns i.e. $E^s \doteq (e_{11}, e_{21}, e_{31}, e_{12}, \cdots, e_{33})^T$. Our goal is to determine this vector and obtain $\mathbf{E}$ by its "un-stacking".

Second, we make use of *Kroneker product* for two image vectors in homogeneous coordinates

$$\mathbf{a} \doteq \mathbf{x}_1 \otimes \mathbf{x}_2 = (x_1 \mathbf{x}_2, x_2 \mathbf{x}_2, 1 \cdot \mathbf{x}_2)^T \,, \tag{3.23}$$

to express the epipolar constrain per 1 correspondence (point) as

$$\mathbf{a}^T E^s = 0 \,. \tag{3.24}$$

Having a set of corresponding image points $\left(\mathbf{x'}_1^{\,i}, \mathbf{x'}_2^{\,i}\right)$, $i = 1, 2, \cdots, n$, we create n vectors $\left(\mathbf{a}^i\right)^T$ and put them into a matrix $\chi \doteq (\mathbf{a}^1; \mathbf{a}^2; \cdots; \mathbf{a}^n)$.

[1] $[\mathbf{t}\times] = \begin{pmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{pmatrix}$

With that we express the epipolar conditions for all correspondences in a system of linear equations

$$\chi E^s = 0\,. \tag{3.25}$$

The solution of this equation is unique if the rank of the matrix χ is exactly 8 (global scale factor cannot be determined). For this reason we need $n \geq 8$ points. Note that expression (3.25) holds only in the absence of noise. However, in reality we have to deal with noise and we are likely to have more correspondences. In the 8-point algorithm the choice is made to minimize the least-square error function of *misclosures* $\|\chi E^s\|^2 \neq 0$, which is achieved by choosing E^s to be an eigenvector of $\left(\chi^T\chi\right)$ that corresponds to its smallest singular value[2] λ. Practically, this can be found by performing a singular value decomposition of $\chi = \mathbf{U}_\chi\Sigma_\chi\mathbf{V}_\chi^T$; i.e. factoring χ into a product of diagonal matrix Σ_χ containing the eigenvalues and orthogonal[3] matrices $\mathbf{U}_\chi$ and $\mathbf{V}_\chi$; and defining E^s to be the column of $\mathbf{V}_\chi$ associated with the smallest singular value. Then we reshape the nine elements of E^s into 3×3 matrix $\mathbf{E}$.

While the reconstructed $\mathbf{E}$ minimizes the norm $\|\chi E^s\|^2$ in the least-square sense, it is not guaranteed – due to the observation of unmodelled errors – that its structure belongs to the space of essential matrices. This space is characterized by $\mathbf{E} = \mathbf{U}\,\mathrm{diag}\{\sigma,\sigma,0\}\mathbf{V}^T$, where $\sigma = \| \mathbf{t} \|$. A common approach is therefore to re-project the estimated $\mathbf{E}$ to such space. This is achieved by carrying the singular value decomposition of $\mathbf{E}$

$$\mathbf{E} = \mathbf{U}\,\mathrm{diag}\left\{\lambda_1,\lambda_2,\lambda_3\right\}\mathbf{V}^T \tag{3.26}$$

with $\lambda_1 > \lambda_2 > \lambda_3 \neq 0$ and then setting the smallest eigenvalue to zero and other two as $0.5\,(\lambda_1 + \lambda_2)$. Alternatively, as the global scale cannot be recovered by image observations only, it may be well chosen as unity, which corresponds to *normalized essential space* where the two largest eigenvalues are set to one.

3.6.3 Pose reconstruction

Lets define a rotation matrix

$$\mathbf{R}_z\left(\pm\pi/2\right) = \begin{pmatrix} 0 & \mp 1 & 0 \\ \pm 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{3.27}$$

where the meaning of $\pm$ and $\mp$ signs will be explained later. Considering that (as for any rotation matrix) $\mathbf{R}_z\mathbf{R}_z^T = \mathbf{I}$ together with the elements

[2] eigenvalue

[3] i.e. $\mathbf{U}^T\mathbf{U} = \mathbf{I}_9$ $\mathbf{V}^T\mathbf{V} = \mathbf{V}\mathbf{V}^T = \mathbf{I}_9$

of singular value decomposition of $\mathbf{E}$, we can verify the correctness of the following relation

$$\mathbf{E} = [\mathbf{t}\times]\,\mathbf{R} = \mathbf{U}\Sigma\mathbf{V}^T = \mathbf{U}\mathbf{R}_z\Sigma\mathbf{R}_z^T\mathbf{V}^T = \underbrace{\mathbf{U}\mathbf{R}_\mathbf{z}\Sigma\mathbf{U}^T}_{[\mathbf{t}\times]}\underbrace{\mathbf{U}\mathbf{R}_z^T\mathbf{V}^T}_{\mathbf{R}}, \quad (3.28)$$

since $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ due to orthonormality of $\mathbf{U}$ after re-projection of $\mathbf{E}$. Then the relative rotation follows directly from (3.28) as the product of three rotation matrices on the right, i.e.

$$\mathbf{R} = \mathbf{U}\mathbf{R}_z^T(\pm\pi/2)\mathbf{V}^T\,, \quad (3.29)$$

and the relative translation (up to a scale) as

$$[\mathbf{t}\times] = \mathbf{U}\mathbf{R}_z^T(\pm\pi/2)\Sigma\mathbf{U}^T\,, \quad (3.30)$$

where it can be proven that $\mathbf{U}\mathbf{R}_z^T(\pm\pi/2)\Sigma\mathbf{U}^T$ is of a skew-symmetric form. The $\pm$ sign in $\mathbf{R}_z(\pm\pi/2)$ reflects the fact that each essential matrix gives two possible solutions and its reconstructed sign is arbitrary. Hence, we could possibly obtain up to four solutions of the relative pose $(\mathbf{R}, \mathbf{t})$ from $\pm\mathbf{E}$. For three of them, either $\mathbf{R}$ is not a rotation matrix ($det\mathbf{R} = -1$) or the re-projected points are not in front of both lens, which is physically not possible. In other words, if all points fall behind both cameras, the translation vector t must be multiplied by -1. We should also mention that despite its simplicity, the 8-point algorithm is not without potential numerical weaknesses that may become apparent in a particular geometry and observation noise. However, as demonstrated by Hartley (2012), these can be avoided by data pre-processing (translation and scaling).

3.6.4 Structure reconstruction

Having 8 or more correspondences as an input, the previously described algorithm determined the relative rotation and translation between the two cameras, the latter up to a global scale (ξ). Setting the norm of translation vector to unity is equivalent of choosing $\xi = 1$. The relative pose can then be used to retrieve the position of the other correspondences on the images in 3-D.

Considering again the relation (3.19) that relates camera poses to n image correspondences

$$\mu_2^i\mathbf{x}_2^i = \mu_1^i\mathbf{R}\mathbf{x}_1^i + \xi\mathbf{t}\,, i = 1, 2, \cdots, n. \quad (3.31)$$

Since $(\mathbf{R}, \mathbf{t})$ are known, this relation is linear and therefore can be easily solved once the unknowns depth μ_1, μ_2 with respect to the first and second camera frames are determined. One of them is, however, redundant, as it is function of $(\mathbf{R}, \mathbf{t})$ as well as the arbitrary choice of the global scale ξ. Hence

we can eliminate, for instance μ_2 by multiplying the above equation by the orthogonal operator $[\mathbf{x}_2\times]$ to obtain

$$\mu_1^i \left[\mathbf{x}_2^i\times\right] \mathbf{R}\mathbf{x}_1^i + \xi \left[\mathbf{x}_2^i\times\right] \mathbf{t} = 0\,. \tag{3.32}$$

An equivalent form that regroups the unknowns in a common vector is

$$\left(\begin{array}{cc} [\mathbf{x}_2^i\times]\,\mathbf{R}\mathbf{x}_1^i, & [\mathbf{x}_2^i\times]\,\mathbf{t} \end{array} \right) \left(\begin{array}{c} \mu_1^i \\ \xi \end{array} \right) \doteq \mathbf{M}^i\mu^{\mathbf{i}} = 0\,. \tag{3.33}$$

To obtain a unique solution the matrix $\mathbf{M}$ needs to be of rank 1, or $[\mathbf{x}_2^i\times]\,\mathbf{t} \neq 0$. Notice, that this is not the case when the point p lies on the line connecting two optical centers.

Regrouping all n correspondences into one equation while noticing that ξ is common to all of them we obtain $\overline{\mu} = \left(\mu_1^1, \mu_1^2, \cdots, \mu_1^n, \xi\right)^T$ and a matrix $\mathbf{M}$ defined as

$$\mathbf{M} \doteq \left(\begin{array}{cccc|c} [\mathbf{x}_2^1\times]\,\mathbf{R}\mathbf{x}_1^1 & 0 & 0 & [\mathbf{x}_2^1\times]\,\mathbf{t} & \\ 0 & [\mathbf{x}_2^2\times]\,\mathbf{R}\mathbf{x}_1^2 & 0 & [\mathbf{x}_2^2\times]\,\mathbf{t} & \\ 0 & 0 & \ddots & 0 & \vdots \\ 0 & 0 & [\mathbf{x}_2^n\times]\,\mathbf{R}\mathbf{x}_1^n & [\mathbf{x}_2^n\times]\,\mathbf{t} & \end{array} \right). \tag{3.34}$$

The solution to the equation

$$\mathbf{M}\overline{\mu} = 0 \tag{3.35}$$

determines all the unknowns in the vector $\overline{\mu}$ up to the last one corresponding to the one global scale ξ. Similarly to the approach of essential matrix determination, the minimization of the square of misclosures (3.35) can be found as the eigenvector of $\mathbf{M}^T\mathbf{M}$ that corresponds to the smallest singular value.

3.6.5 Global scale

The global scale can be determined only by some exterior knowledge either on the camera motion, as discussed in Sec. 4.4, or on the object coordinates of some observed points. For instance, if the restitution of structure is required in a mapping frame we need to know the coordinates of at least 3 points in both frames to apply the 7 parameter similarity transformation. Finally, the problem of reconstruction can be formulated as unconstrained optimization problem, where the minimization is searched with respect to all unknowns $\mathbf{x}_1^i, \mathbf{R}, \mathbf{t}, \bar{\mu}$. This is known in the literature as bundle adjustment, and its method will be further detailed in the next section as well as in Sec. 4.3. The presented form, however, allows to develop the needed approximations for its effective solution that are based exclusively on image observations.

3.7 Scene from multiple views

3.7.1 Multiple-view matrix

We now consider the existence of more than two views of the same object, which is rather a standard case. Without the lost of generality, lets take the frame of the first camera as a reference frame for 3-D reconstruction. With m views/images at disposition we obtain from (3.17) the following projection matrices

$$\Pi_1 = (\mathbf{I}, \mathbf{0}) \, , \Pi_2 = (\mathbf{R}_2, \mathbf{t}_2) \cdots \Pi_m = (\mathbf{R}_m, \mathbf{t}_m) \, , \tag{3.36}$$

Considering at the moment only one point p and applying a similar development as for relations (3.31) - (3.33) we can derive the *multiple-view matrix* $\mathbf{M}_p$. We do so, by inserting into two columns of $\mathbf{M}_p$ a coplanarity constrain (3.19 of view i between the first and ith camera reference frame. Up to its depth this constrain is $\mu_1 \left[\mathbf{x}_i \times\right] \mathbf{R}_i \mathbf{x}_1 + \left[\mathbf{x}_i \times\right] \mathbf{t}_i = 0$. In a matrix form this is

$$\mathbf{M}_p \begin{pmatrix} \mu_1 \\ 1 \end{pmatrix} = 0 \, , \tag{3.37}$$

with $\mathbf{M}_p$ defined as

$$\mathbf{M}_p \doteq \begin{pmatrix} \left[\mathbf{x}_2 \times\right] \mathbf{R}_2 \mathbf{x}_1 & \left[\mathbf{x}_2 \times\right] \mathbf{t}_2 \\ \left[\mathbf{x}_3 \times\right] \mathbf{R}_3 \mathbf{x}_1 & \left[\mathbf{x}_3 \times\right] \mathbf{t}_3 \\ \vdots & \vdots \\ \left[\mathbf{x}_m \times\right] \mathbf{R}_m \mathbf{x}_1 & \left[\mathbf{x}_m \times\right] \mathbf{t}_m \end{pmatrix} \tag{3.38}$$

This matrix thus associates m views of point p by involving both the image $\mathbf{x}_1$ and the co-images $\left[\mathbf{x}_2 \times\right], \left[\mathbf{x}_3 \times\right], \cdots, \left[\mathbf{x}_m \times\right]$. In other words, it encodes all constraints that exists among the m views of a point. It has a rank 1, as long as the pair of vectors $\left[\mathbf{x}_i \times\right] \mathbf{t}_i$, $\left[\mathbf{x}_i \times\right] \mathbf{R}_i \mathbf{x}_i$ is linearly dependent for each $i = 1, 2, \cdots, m$, which is equivalent to the *bilinear epipolar constraints*.

$$\mathbf{x}_i^T \left[\mathbf{t}_i \times\right] \mathbf{R}_i \mathbf{x}_1 = 0 \, . \tag{3.39}$$

In such a situation the projection of p on the image is unique, which is not the case for $\mathtt{rank}\left(\mathbf{M}_p\right) = 2$ or $\mathtt{rank}\left(\mathbf{M}_p\right) = 0$. Rank testing can be potentially used for filtering out the wrongly established correspondences in feature matching.

3.7.2 Trilinear constraint

In some situations it maybe useful to formulate one condition involving directly three-views. Let's consider one point that is viewed by three cameras

$1, i, j$. For this situation we can write two separate coplanarity constrains, the second being transposed

$$\begin{aligned} \mu_1 \left[\mathbf{x}_i\times\right] \mathbf{R}_i \mathbf{x}_1 &= -\left[\mathbf{x}_i\times\right] \mathbf{t}_i \\ \mathbf{x}_1^T \mathbf{R}_j^T \left[\mathbf{x}_j\times\right]^T \mu_1 &= -\mathbf{t}_j^T \left[\mathbf{x}_j\times\right]^T \end{aligned} \tag{3.40}$$

Multiplying across the left- and right-hand sides of both equations (3.40) and making them equal

$$-\left[\mathbf{x}_i\times\right] \mathbf{R}_i \mathbf{x}_1 \mathbf{t}_j^T \left[\mathbf{x}_j\times\right]^T = \left[\mathbf{x}_i\times\right] \mathbf{t}_i \mathbf{x}_1^T \mathbf{R}_j^T \left[\mathbf{x}_j\times\right]^T , \tag{3.41}$$

then rearranging the terms to one side we obtain the *trilinear constraint*

$$\left[\mathbf{x}_i\times\right] \left(\mathbf{t}_i \mathbf{x}_1^T \mathbf{R}_j^T - \mathbf{R}_i \mathbf{x}_1 \mathbf{t}_j^T\right) \left[\mathbf{x}_j\times\right] = 0 . \tag{3.42}$$

The trilinear constrain implies a bilinear constrains (3.39), except for a special case in which $\left[\mathbf{x}_j\times\right] \mathbf{t}_j = \left[\mathbf{x}_j\times\right] \mathbf{R}_j \mathbf{x}_i = 0$ for some view j. In this rare situation the point p lies on the line connecting the optical centers $\mathbf{o}_1, \mathbf{o}_j$. The application of trilinear constraint may therefore be of a certain advantage for some special cases, such as that when three image vectors of the same point are coplanar.[4] When they still satisfy the trilinear constrain, 3-D coordinates of this point can be reconstructed. It should be also mentioned that any other algebraic constraint among m images can be reduced to those involving either two or three at a time (i.e. application of either bilinear or trilinear constraints).

3.7.3 Processing strategies

The processing strategies for handling multiple views vary in function of image geometry, scene texture (goodness of feature detection, matching and filtering), camera calibration, data noise and experience. We present therefore only the main concepts, while leaving the details of their combination into to a particular implementation. These are schematically depicted in Fig. 3.6.

In principle, any multiple-view can be broken down into a sequence of two-view scenarios between first and last camera poses. This situation is highlighted in the upper part of the Fig. 3.6 and is often used in practice when the overlap between images is small, the texture allows finding only few correspondences or there is a large uncertainty in the camera model. To mitigate the accumulation of random influences in the sequence, the "two view step" is followed by a global optimization involving all views.

The second approach is to use 8-point algorithm only once for some initial pair of view and perform global optimization on the rest as showed in

[4]This may be the case, for instance, in car-based mapping system when views from the same forward looking camera are combined between successive times, i.e. involving displacement only along the depth of field.

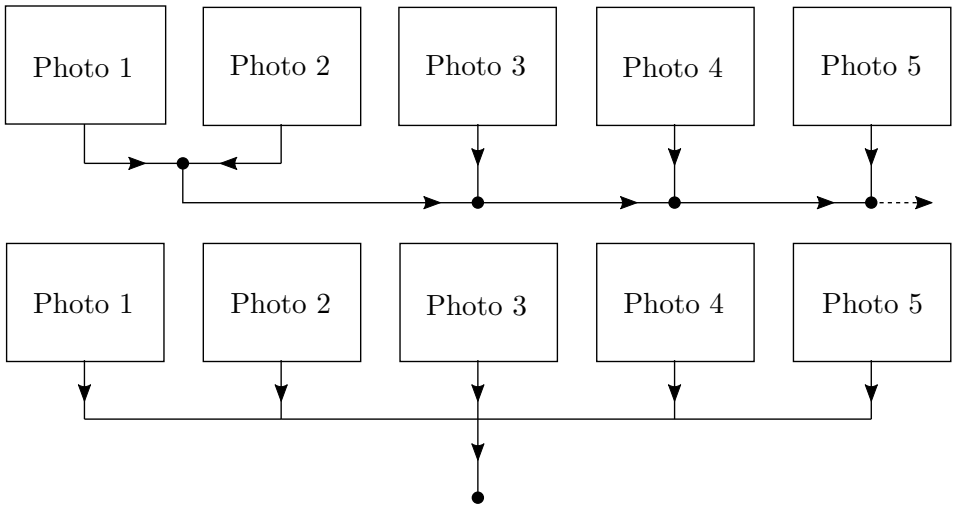


Figure 3.6: Image reconstruction strategies: incremental (top) versus global (bottom).

the bottom part of Fig. 3.6. This method is more suitable for good image geometry, large overlap and (pre)calibrated camera. It involves constructing relation containing multiple view matrix $\mathbf{P}_i$, similar to that of (3.38), but involving m images $\mathbf{x}_1^j, \mathbf{x}_2^j, \cdots, \mathbf{x}_1^m$ of n points p^j, $j = 1, 2, \cdots n$ from which we would like to estimate the unknown projections $\Pi_i \left(\mathbf{R}_i^s, \mathbf{t}_i\right)^T$, $i = 2, 3, \cdots, m$

$$\mathbf{P}_i \begin{pmatrix} \mathbf{R}_i^s \\ \mathbf{t}_i \end{pmatrix} \begin{pmatrix} \left[\mathbf{x}_1^1\times\right]^T \otimes \left[\mathbf{x}_i^1\times\right] & \lambda^1 \left[\mathbf{x}_i^1\times\right] \\ \left[\mathbf{x}_1^2\times\right]^T \otimes \left[\mathbf{x}_i^2\times\right] & \lambda^2 \left[\mathbf{x}_i^2\times\right] \\ \vdots & \vdots \\ \left[\mathbf{x}_1^n\times\right]^T \otimes \left[\mathbf{x}_i^n\times\right] & \lambda^n \left[\mathbf{x}_i^n\times\right] \end{pmatrix} \begin{pmatrix} \mathbf{R}_i^s \\ \mathbf{t}_i \end{pmatrix} = 0\,, \tag{3.43}$$

where $\otimes$ is the *Kronecker product* between matrices [5], λ^j is the inverse of unknown depth μ^i. The matrix $\mathbf{P}_i$ is of size $3n \times 12$ and is of rank 11 if more than $n \geq 6$ points are provided. Then the projection matrix $\Pi_i = (\mathbf{R}_i, \mathbf{t}_i)$ can be solved for up to the scale factor. When $\mathbf{P}_i$ is of rank higher than 11, the solution that minimizes the square of misclosures is obtained as an eigenvector of $\mathbf{P}_i$ associated with the largest eigenvalue. As the estimate of Π_i is affected by random errors, the estimated matrix $\mathbf{R}_i$ need to be re-projected to the rotational space $SO\,(3)$ and the vector $\mathbf{t}_i$ re-scaled. Assuming the pose for the second view is found by the 8-point algorithm, the scalars λ^j can be determined from the first row of (3.43)

[5] Similar to that of (3.23) but with the elements of the left matrix stacked into a vector.

involving $\lambda^j \left[\mathbf{x}_2^j \times\right] \mathbf{t}_2 = -\left[\mathbf{x}_2^j \times\right] \mathbf{R}_2 \mathbf{x}_1^j$. These initial value of λ^j can then be used for the recovery of Π_i, $i = 3, 4, \cdots, m$. Since $\mathbf{t}_2$ is recovered from the 8-point algorithm up to a scale factor, the other views are recovered from that up to a global single scale.

Optimization

The equivalent and perhaps even simpler formulation of the global optimization, i.e. the concurrent determination of object coordinates, camera parameters and pose (i.e. structure and motion) is presented under a name of *bundle adjustment*, which extension that accommodates also other inputs is presented in Sec 4 on Sensor Fusion. Bundle adjustment received its name after application of ray-tracing collinearity condition (3.17) $\mu\overline{\mathbf{x}'} = \Pi\overline{\mathbf{x}}^m$ on a "bundle of rays" connecting object points with its projection on the image in combination with a particular sensor model (3.18). Nevertheless, this optimization approach requires linearization and that an existence of approximate values of parameters. With an exclusive use of image observations, the approximate value of parameters can be obtained by the methods described in this and previous sections.

3.8 Feature matching

The term image matching stands for the mostly automatic reference between regions or pixels of two or more images that represent the same feature or point in the object space. Automatic aerial triangulation (AAT) requires the availability of suitable image-matching tools as a key component. These tools should enable fully automatic tie-point measurement by providing homologous features with suitable accuracy and reliability. For this purpose, feature-based matching approaches are frequently used. First, primitives suitable for image matching are extracted, while in a second step their correspondences are determined by some similarity and consistency measures. These two steps of feature-based matching techniques result in a categorization into feature detectors and feature descriptors. Detectors search for image points or regions which are geometrically stable under different transformations and that contain high information content. The results are generally called interest points, corners, or invariant regions. Descriptors instead analyze the image to provide a 2-D vector of image information at those areas defined by the respective interest point. The subsequent matching process then exploits this information for similarity measurement in order to evaluate potential point correspondences. To remove outliers remaining after this matching, geometric constraints such as epipolar geometry are applied by robust estimators in a final step.

Feature extraction and matching are strongly related; however, these two steps are discussed separately in the next sections. This separation also re-

sults from the high accuracy demands within automatic aerial triangulation, which is usually fulfilled by hybrid matching approaches. In this context, tie-point positions as provided by feature-based methods are refined in a subsequent step using intensity-based correlation strategies.

3.8.1 Image matching primitives

To detect primitives suitable for image matching, so-called interest operators were first developed in the 1970s. Since then, a wide variety of algorithms have evolved in computer vision, pattern recognition, and photogrammetry. Comprehensive overviews on feature extraction are given, e.g., in Schmid et al (2000); Jazayeri and Fraser (2010).

In the context of image matching, feature extraction aims to identify primitives, which are invariant against radiometric and geometric distortions, robust against image noise, and distinguishable from other points (Haralick and Shapiro, 1992). This task is especially complex for close-range applications, in which one frequently has to cope with convergent images with different look angles at varying scale. However, the situation is easier for aerial triangulation. In this context, similar viewpoints and relatively short time intervals during image collection avoid problems due to perspective distortions and large changes in illumination. Furthermore, matching can be simplified using a priori information on the respective image geometry, which is usually available from camera calibration, the standardized flight geometry of airborne image blocks, or measured GNSS trajectories. Within commercially available AAT software, the Förstner operator (W. Förstner, 1987) has been widely used. This operator was developed for fast detection and precise location of distinct points including corners and centers of circular image features.

Feature detectors such as the Förstner and Harris operators were mainly integrated for applications in airborne photogrammetry (Harris and Stephens, 1988). Meanwhile, the scale-invariant feature transform (SIFT) key point detector (Lowe, 2004) has become the quasi-standard for point extraction and matching. Scale-invariant means that a feature in object space that appears with a large scale in one image and with a small scale on the other still can be detected as the same by the SIFT-operator. It is scale invariant since feature points are detected in the so-called scale space by searching for maxima in an image pyramid as defined by a stack of the difference of Gaussians (DoG) (Lowe, 2004). Thus, it became especially popular in close-range applications, where matching is frequently aggravated due to the appearance of larger perspective distortions.

3.8.2 Feature matching strategies

Feature detection is followed by a suitable matching step to provide the required point correspondences for the aspired AAT. This matching is based on information representing the local image patch in the vicinity of the respective feature point. Attributes are usually derived from the gray or gradient values in the features neighborhood. As an example, the feature description for the SIFT operator is generated from the histogram of the gradient vectors in the local neighborhood of the key point location Lowe (2004). This approach transforms the image data into a scale- and rotation-invariant representation. A pair of key points within two overlapping images is then regarded as corresponding if the Euclidian distance between their respective descriptors is less than a given threshold and the distance to the second nearest descriptor is greater than a second given threshold. An overview on the use of local descriptors is given in Mikolajczyk and Schmid (2005).

If feature matching is required during the evaluation of aerial imagery, the homogeneity conditions during image collection usually allow for the use of gray values in the local vicinity of a feature point. Thus, the similarities of potentially corresponding image patches can be measured by normalized cross-correlation (NCC)

$$\rho(r,c) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}[g_L(i,j) - \bar{g}_L] \cdot [g_R(r+i, c+j) - \bar{g}_R]}{\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}[g_L(i,j) - \bar{g}_L]^2 \cdot \sum_{i=1}^{m}\sum_{j=1}^{n}[g_R(r+i, r+j) - \bar{g}_R]^2}} \tag{3.44}$$

where

ρ	normalized cross correlation
r, c	row and column
m, n	shift of rows and columns between both images
g_L, g_R	gray-values of a pixel in left and right image
$\bar{g}_L, \bar{g}_R$	average gray-values of search window in left and right image

This provides values normalized in the interval with highest similarity for a coefficient close to 1. Usually, such similarity measurements are not robust enough to avoid mismatches. Hence, an additional step to reject potential outliers is required. For this purpose, geometric constraints as provided from epipolar geometry of the respective image pair are frequently used. As an example, algorithms based on random sample consensus (RANSAC) (Fischler and Bolles, 1981) robustly estimate the relative orientation between image pairs.

This provides a suitable transformation for the corresponding image points and therefore allows elimination of potential mismatches while providing consistent point correspondences. The algorithm can be summarized as follows.

1. A random sample of five correspondent corresponding image points is taken from the list of matched points of the two images.

2. From these five correspondences, the relative orientation of the image pair is computed using the algorithm described in Nistér (2004) or, alternatively, with (3.28) when setting the translation vector to unity.

3. This relative orientation defines for each image point $j = 1, 2, \cdots, n$ in the left image the corresponding epipolar line in the right image (3.22), $\mathbf{x}_2^j \mathbf{E} \mathbf{x}_1^j = 0$. The difference between this epipolar line and the corresponding point in the right image defines an error for this potential match with respect to the calculated relative orientation. If a matched point pair has a small epipolar error, it fits well with the estimated relative orientation. In that case this potential correspondence is considered as a hypothetical inlier, otherwise it is an outlier as schematically depicted in Fig. 3.7.

4. If sufficiently many point pairs are classified as inliers, the estimated relative orientation is reasonably good and the algorithm can be terminated. All inliers are preserved, while the outliers are eliminated from the final list of correspondences.

5. Otherwise, the RANSAC algorithm continues with step 1 with another random sample of five correspondent image points.

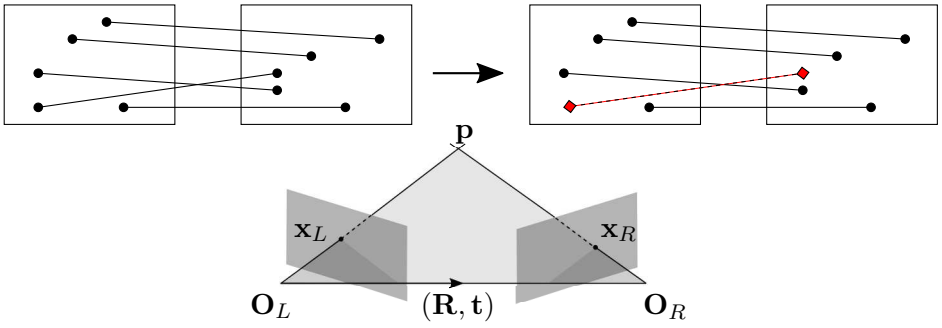


Figure 3.7: Filtering of wrongly asigned correspondences on randomly selected subset of points through the epipolar constraint.

These correspondences can be directly used as tie-points during optimisation with other data. However, the accuracy of the applied feature-based matching is usually increased to subpixel level by subsequent area-based matching. This can, e.g. be realized using the NCC as defined by (3.44). For subpixel measurement, the center of the correlation masks g_L and g_R are again defined by the coordinates of the left and right feature points.

The NCC is then computed in a local 3 × 3 neighborhood of the potential match. Of course, the best similarity position defined by the maximal NCC coefficient will correspond to the center point of this matrix, i.e., the coordinates of the corresponding right feature point. However, the correlation coefficients in the local neighborhood of this best match position can be used for subpixel refinement by interpolation through the second-order polynomial. The cross-sections in row- and column-direction are parabolas.

$$f(r,c) = a_0 + a_1 r + a_2 c + a_3 r^2 + a_4 rc + a_5 c^2. \tag{3.45}$$

$$\begin{pmatrix} \rho_0(-1,-1) & \rho_1(-1,0) & \rho_2(-1,1) \\ \rho_3(0,-1) & \rho_4(0,0) & \rho_5(0,1) \\ \rho_6(1,-1) & \rho_7(1,0) & \rho_8(1,1) \end{pmatrix} \tag{3.46}$$

which represents the NCC coefficients for a × local neighborhood centered at position $(0,0)$ of the maximum value p_4. The NCC coefficients $\mathbf{l} = (\rho_0, \rho_1, ..., \rho_8)^T$ computed by (3.44) for the different positions (r_i, c_i) are then used as observations within the Gauss-Markov model $\mathbf{Ax} - \mathbf{l} = \mathbf{v}$. The parameters of the polynomial (3.45) can then be estimated with

$$\mathbf{A} = \begin{pmatrix} 1 & r_0 & c_0 & r_0^2 & r_0 c_0 & c_0^2 \\ 1 & r_1 & c_1 & r_1^2 & r_1 c_1 & c_1^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & r_8 & c_8 & r_8^2 & r_8 c_8 & c_8^2 \end{pmatrix} \tag{3.47}$$

If the available values for (r_i, c_i) are introduced, the Gauss-Markov model results in the equation

$$\begin{pmatrix} 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 0 & 1 & 0 & 0 \\ 1 & -1 & 1 & 1 & -1 & 1 \\ 1 & 0 & -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{pmatrix} - \begin{pmatrix} \rho_0 \\ \rho_1 \\ \rho_2 \\ \rho_3 \\ \rho_4 \\ \rho_5 \\ \rho_6 \\ \rho_7 \\ \rho_8 \end{pmatrix} = V \tag{3.48}$$

The standard solution

$$\mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{l} = [\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}]^T\mathbf{l} = \mathbf{t}^T \cdot \mathbf{l},$$

then provides the five parameters of the polynomial $\mathbf{x} = (a_0, a_1, ..., a_4)^T$. The partial derivatives then define the extremum of this polynomial in row- and column-direction by

$$\frac{\partial f}{\partial r} = a_1 + 2a_3\Delta r + a_4\Delta c \cong 0 \tag{3.49}$$

$$\frac{\partial f}{\partial c} = a_2 + a_4\Delta r + a_5\Delta c \cong 0 \tag{3.50}$$

This finally gives the subpixel refinement $(\Delta r, \Delta c)$ for the initial center position of the best match as

$$\Delta r = \frac{a_2 a_4 - 2a_1 a_5}{4a_3 a_5 - a_4^2} \tag{3.51}$$

$$\Delta c = \frac{a_1 a_4 - 2a_2 a_3}{4a_3 a_5 - a_4^2} \tag{3.52}$$

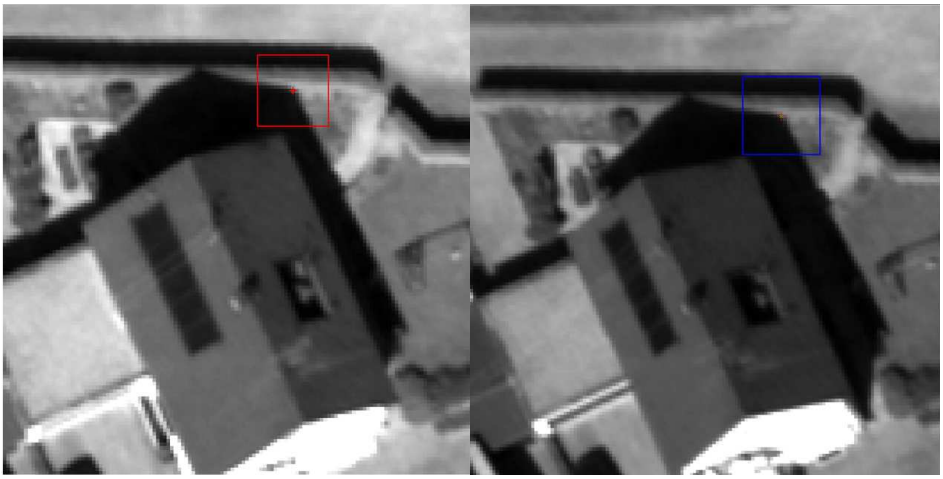


Figure 3.8: Feature point with correlation mask (left image) and search mask (right image).

An example input for the computation of subpixel refinement can be found in Fig. 3.8, which depicts a correlation mask in the left image and a search mask in the right image, both centered at their corresponding feature point. The NCC coefficients ρ_i computed from using the correlation mask within the 3×3 neighborhood of the right feature point give the matrix

$$\begin{pmatrix} 0.9300 & 0.9088 & 0.8622 \\ 0.9862 & 0.9922 & 0.9664 \\ 0.9281 & 0.9646 & 0.9696 \end{pmatrix}$$

If these values are used as input for (3.48) and (3.51), this gives a shift of $\Delta r = -0.0409$ pixels and $\Delta c = 0.2366$ pixels.

The use of local sample points to determine the interpolated location of the maximum has also been used in the context of SIFT key point extraction (Lowe, 2004). In this application, pixel coordinates x, y and scale s define a 3D scale space function $D(x, y, s)$. Thus subpixel and subscale coordinates $\mathbf{x} = (x; y; s)$ of a feature are found by interpolation with a 3D quadratic function which has the shape of a parabola in each of the three dimensions. This provides a substantial improvement to matching and stability.

As an alternative to NCC-based subpixel measurement of tie-point coordinates, least-squares image matching (W. Förstner, 1987) can be applied. This approach estimates the geometric and radiometric transformations between corresponding patches g_L and g_R from the left and right image, respectively, using the observation equation

$$g_L(r, c) + v = h_1 + h_2 g_R[(a_0 + a_1 r + a_2 c), (b_0 + b_1 r + b_2 c)] \tag{3.53}$$

This approach models geometric differences between image patches by a simple affine transformation with parameters a_0, a_1, a_2, b_0, b_1, and b_2, while radiometric differences caused, e.g. by different sun lighting are represented by offset and gain, h_1 and h_2. The transformation parameters are then estimated through iterate least-squares adjustment. Eq. (3.53) equals the Gauss-Markov model, which minimizes the squared sum of errors of all the observations $v^T v \rightarrow min$. Gray value differences between the corresponding image patches are used for a typical window size of 15×15 pixel, resulting in matching precision of 0.1 – 0.01 pixels.

3.8.3 Dense matching

Stereo-matching aiming at the automatic generation of elevation data from aerial images was already introduced more than two decades ago. Originally, feature based algorithms were applied to extract feature points and then search the corresponding features in the overlapping images. The restriction to matches of selected points usually provides correspondences at high certainty. However, feature based matching was also introduced to avoid problems due to limited computational resources. In contrast, recent stereo algorithms aim on dense, pixel-wise matches. By these means 3D point clouds and Digital Surface Models (DSM) are generated at a resolution, which corresponds to the ground sampling distance GSD of the original images. To compute pixel matches even for regions with very limited texture, additional constraints are required. Local or window based algorithms like correlation use an implicit assumption of surface smoothness since they compute a constant parallax for a window with a certain number of pixels. Those local algorithms establish references between images only under consideration of the gray-value properties of a small environment. This may be error-prone, because small variations of the gray-values and repetitive patterns are difficult to control. In contrast, so-called global algorithms use an

explicit formulation of this smoothness assumption, which is then solved a global optimization problem (Szeliski, 2010). Those global algorithms allow in a way a comparison of the results of the local computation and thus allow for a detection of mismatches and the subsequent deletion of the outliers. One example is scanline optimization, which can be solved very efficiently by recursive algorithms. The scanline optimization is applied where beforehand the local image analysis has been done row by row with mostly a not fully fitting edge detection. The resulting image looks frayed. Using the scanline optimization, the edge points are averages with a geometrically proper and nicely looking result. A very popular and well performing example is semi-global matching (Hirschmüller, 2008), which evaluates a cumulative cost function from the scanlines in the 8 cardinal directions East, Northeast, North etc.. Though the algorithm operates in two dimensions, it is still fast, because it substitutes the 2D-computation by 8 1D computations. Especially when combined with sophisticated aggregation strategies it can produce accurate results very efficiently (Szeliski, 2010).

The progress of software tools for image based DSM generation is also documented by a benchmark conducted by the European Spatial Data Research Organization (EuroSDR) (Haala, 2014). This benchmark tested the DSM programs listed in Tab. 3.8.3.

Name of Software	Manufacturer	Location
SocetSet 5.6 (NGATE)	BAE Systems	Newcastle–Tyne, UK
UltraMap V3.1	Microsoft, Vexcel	Graz, Austria
Match-T DSM 5.5	Trimble/inpho	Stuttgart, Germany
ImageStation ISAE	Geosystems GmbH	Munich, Germany
Pixel Factory	Astrium GEO-Information Services	Paris, France
RMA DSM Tool	Royal Military Academy (RMA)	Brussels, Belgium
Remote Sens. software	Joanneum Research	Graz, Austria
MicMac	IGN France	Paris, France
SURE	IfP, Univ. Stuttgart	Stuttgart, Germany
SGM - FPGA ver.	German Aerospace Centre DLR	Oberpfaffenhofen, Germany
XProSGM	Leica	Heerbrugg, Switzerland

Table 3.3: EuroSDR benchmark on image matching (Haala, 2014).

In addition to the generation of 2.5D models like DSMs and DTMs, the extraction of real 3D structures especially in dense and complex urban

environments is getting more and more important. Such meshed surface representations are widely used for visualization purposes. Moreover, since they directly represent neighborhood information they are especially useful in follow up processes aiming on semantic interpretation of 3D data.

Chapter 4
Sensor Fusion

4.1 Principle

In Sec. 3 we described how to reconstruct a 3-D scene using structure-from-motion techniques. Such method rely solely on image observations to determine the relative orientation between images at an arbitrary scale (e.g., one). Other observations needs to be added to resolve the scale correctly and to obtain the coordinates of objects in a reference/mapping frame. Ideally, such additional information is fused together with image observations in a way that allows an optimal recovery of all involved parameters included those related to the unknown parameters of optical sensor for the purpose of mapping. This, generally, requires determining

1. the sensor interior orientation (IO) parameters. In case of frame/line imagery, a basic set of IO parameters may comprise the focal length (f) or principal distance (c), the principal point (x_0, y_0) and lens distortions that allow consistent interpretation of sensor data[1],

2. the sensors absolute exterior orientation[2] (EO) that, similarly to relative orientation, geometrically connects the sensor data among them, plus with respect to the real world (Sec. 2),

3. auxiliary system parameters that geometrically relates different sensors with respect to each other in space and time such as lever-arms, bore-sights or time-stamping offsets.

As depicted in Fig. 4.1, the process of sensor orientation may take different paths (dashed versus full lines):

- The sensor data can be oriented directly using navigation technology (Sec. 2).

- The sensor is oriented indirectly by identifying corresponding features across overlapping parts of data and connecting them with external references on the ground (Sec. 3.7). This is achieved by a procedure that will be referred to as bundle block (or strip) adjustment. In the

[1] in case of Lidar this may be range-finder bias or misalignment between the laser-beam reflecting mirror and its angular encoder

[2] this is also referred to as pose when limited to position and attitude

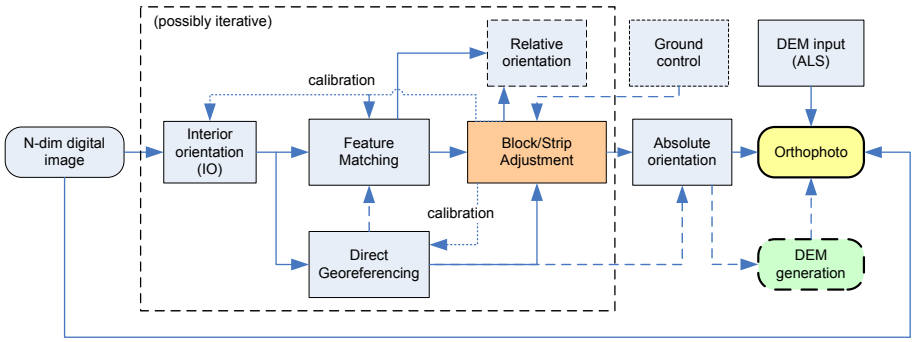


Figure 4.1: General overview of sensor fusion, with the optimalization step denoted as block/strip adjustment, known also as assisted aerotriangulation (AAT).

context of airborne mapping with passive imagery, this approach is called aerial triangulation (AT).

- The methods of direct and indirect sensor orientation can be combined together by extending the adjustment input to block/strip for navigation data. In this case, the procedure is named integrated sensor orientation and can be considered as an extension of the aforementioned block adjustment/AT that is referred to as assisted aero-triangulation (AAT). In robotics, this is operated sequentially and possibly in real time, reason for which it is called simultaneous localization and mapping (SLAM).

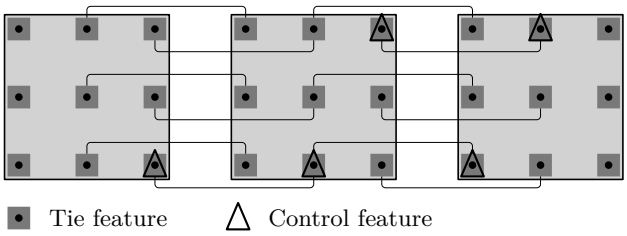


Figure 4.2: Feature (point) conditioning on overlapping imagery.

Aerial triangulation or SLAM belong to the category of network adjustment techniques that make use of redundant information in the overlapping parts of optical data (either pair-wise, strip-wise or block wise; Fig. 4.2). The overlapping segments are called homologous features and range from geometrical primitives as points or lines to more complex features such as surfaces. These features are conditioned within the sensor models to take same coordinates in the object coordinate system. This approach is appli-

cable to passive (Sec. 1.2) as well as active (Sec. 1.4) optical sensors and is indispensable for calibration purposes.

The navigation data usually enter the adjustment as absolute or relative poses, as shown in the upper block of Fig. 4.3. If satellite positionig is absent (indoor environment), intermittent (terrestrial vehicles), or the inertial observations are of poor quality (small UAVs, mobile robots), the trajectory determination process may result in time-varying biases in position and attitude, character of which cannot be correctly modelled within the network of this type. In such situation is better to introduce the original inertial readings (i.e., angular rates and specific forces) directly into a modified network as depicted in the lower block of Fig. 4.3. This approach is rigorous but requires some special care when introducing the differential equations relating the inertial observation to poses. This method proposed by Colomina and Blazquez (2004) under the term dynamic network bears number of advantages as well as challenges. With number of simplifications that are not appropriate for airborne mapping this approach is also employed in robotic indoor SLAM where is reffered to as pose-graph estimation (Strasdat et al, 2010). Its exteded modeling applicable to large-scale mapping project is described in Cucci et al (2017).

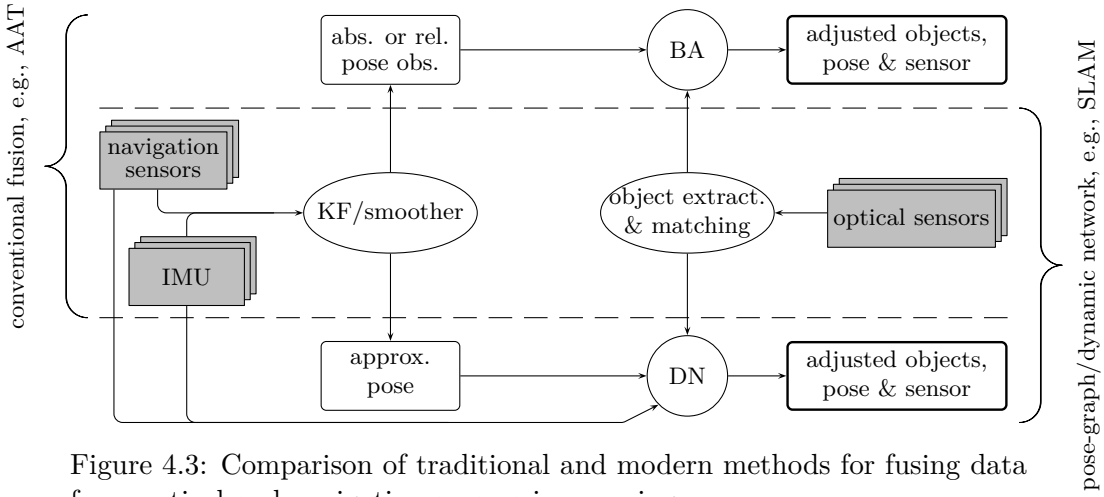


Figure 4.3: Comparison of traditional and modern methods for fusing data from optical and navigation sensors in mapping.

Sensor	**Exterior**	**Interior and system parameters**
Frame cameras	Position, attitude, (constant trajectory errors per block or strip)	Lens distortions, PP, PD, (temperature, pressure, boresight, lever-arm, synchronization, etc.)
Line cameras	Position, attitude, (constant trajectory errors per block or strip)	Lens distortions, PP, PD, (temperature, pressure, boresight, lever-arm, synchronization, etc.)
Lidar	Position, attitude, (constant trajectory errors per block or strip)	Boresight, (range-finder offset, mirror distortion & alignment, encoder scale & offset, etc.)
Radar	Position, attitude, velocity	Doppler, boresight, lever-arm, (synchronization), etc.
High-orbit satellites	Polynomial parameters	(Sensor dependent)
Low-orbit satellites	Position, attitude (or polynomial parameters)	(Sensor dependent)

Table 4.1: Example of parameters for sensor orientation and calibration.

4.2 Parameters

The goal of setting up a network is the optimal fusion of all sensor data to determine concurrently and optimally the coordinates of the image features in the mapping frame together with the set of orientation parameters (exterior, interior, system). An example of parameters sets for different sensors is presented in Tab. 4.1. More specifically, in the example of frame/line cameras, these unknown parameters are

- 3-D positions of distinctive features identified in the images (e.g., tie points), $\mathbf{p}_n^m$, with $n \in \{1; \cdots ; N\}$,
- (optionally), the basic interior orientation parameters, i.e., the camera constant c, the principal point (x_0, y_0),
- (optionally), the additional interior orientation parameters represented either by physical or the replacement models (Sec. 4.3),
- samples of the IMU-body frame pose for each camera exposure j, i.e., the position and the attitude of body frame with respect to mapping frame m; $\Gamma_{b,j}^m = [\mathbf{x}_b^m(j), \mathbf{R}_b^m(j)]$, with $j \in \{1; \cdots ; J\}$,
- (optionally) the body to camera lever-arm and boresight $\mathbf{x}_{bs}^b$ and $\mathbf{R}_s^b$,

- (optionally[3]) GNSS antenna lever-arm $\mathbf{x}_{ba}^{b}$, or $\mathbf{x}_{sa}^{s}$, respectively,
- (optionally[4]) INS systematic errors, e.g. random ,yet time-constant 3-D bias vectors for the gyroscopes $\mathbf{b}_{\omega}$ and accelerometers $\mathbf{b}_{f}$.

The observation models related to interior orientation of some optical sensors and those related to navigation sensors are presented in the following.

4.3 Optical distortion models

Optical distortions directly influence the metric quality of the image and therefore have to be considered. As introduced in Sec. 1.1 for the case of an ideal camera where the incident and emerging nodal points define its optical axes, the chief or central rays pass through the lens without deviation while the emerging ray remains parallel to the original incident ray. The deviation from this ideal, parallel, case needs to be modelled (and later estimated by the calibration setup) since the ideal assumptions cannot be perfectly met in a real camera system design. Such deviations can be best captured by models that relates to the physical properties of the system. When this is not possible either due to the unknown properties of the systems or its high complexity, it may be better to adopt some general models (e.g., polynomial) and determine a subset of relevant parameters.

4.3.1 Sensor physical models

In many frame/line cameras, the symmetric lens distortion have the most relevant influence on 3-D object point reconstruction. Relation (3.18) introduced a basic distortion model of perspective-centred image coordinates. A more general model is the Conrady-Brown distortion correction (Freyer and Brown, 1986) relating the distorted image coordinates (x_d, y_d) to the undistorted (x, y) through third-order radial $[k_1, k_2, k_3]$ and second-order tangential $[p_1, p_2]$ coefficients

$$\begin{aligned} x_d &= x\left(k_1 r^2 + k_2 r^4 + k_3 r^6\right) + p_1\left(r^2 + 2x^2\right) + 2p_2 xy \\ y_d &= y\left(k_1 r^2 + k_2 r^4 + k_3 r^6\right) + p_2\left(r^2 + 2y^2\right) + 2p_1 xy \end{aligned} \tag{4.1}$$

with $r^2 = x^2 + y^2$. Affinity as well as non-orthogonality effects on image coordinates maybe added to (4.1) and the size of a particular calibration set may be even larger depending on the system and the type of calibration (Gruen, 1982). When radial distortion is present, the image point is displaced radially in comparison with its ideal position. If this displacement is positive, i.e., the point is shifted towards the image borders, the distortion is

[3] in dynamic network/SLAM or in AAT if GNSS position is used instead of GNSS/INS
[4] in dynamic network/SLAM

referred to as barrel distortion; if the distortion is negative, it is referred to as pincushion distortion. Radial distortions can be balanced through proper adaptation of the focal length *f*. As can be seen from Fig. 4.4, the distortion remaining after balancing is small. Since this modified value is an outcome of the camera calibration and different from the physical focal length, it is now called the calibrated focal length or camera constant *c*.

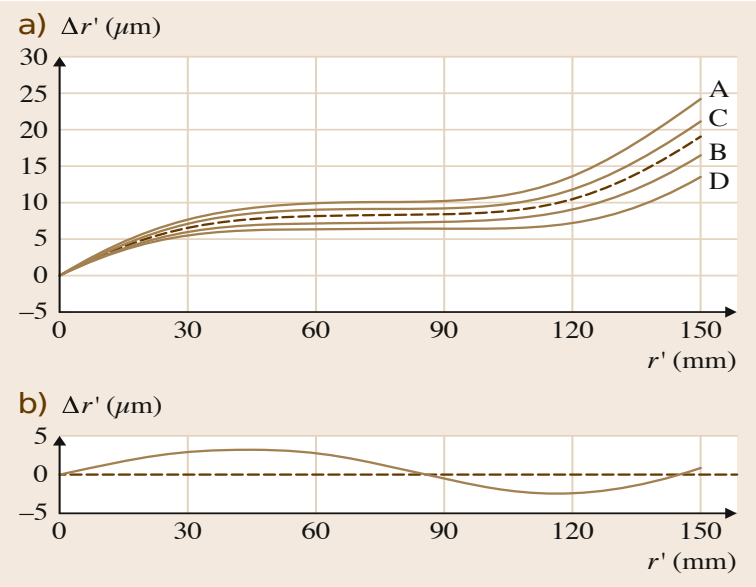


Figure 4.4: (a) Radial distortion curves (A-D) and its mean value (dashed-line). (b) Mean radial distortion after balancing, after Kraus (2007).

4.3.2 Sensor replacement models

A sensor replacement model is a model that approximates the original, or rigorous, sensor model associated with a specific sensor by an arbitrary function. Although such models hide the details of the physical sensor model, they have some advantages, being possibly applicable across different sensors. Also, their evaluation may be faster for obtaining ground-to-image coordinates. This is especially interesting for voluminous satellite data or real-time mapping applications. Examples of replacement sensor models include

- 3-D polynomial model,
- affine line-based transformation model (e.g., satellite imagery),
- rational polynomial coefficients model (e.g., satellite imagery),

- universal sensor model (e.g., in general image-processing packages).

Such replacement models are successfully applied for transfer of sensor orientation to the final users; however, their usage in block adjustment is less appropriate.

4.4 Observation models

4.4.1 Image observations

Frame cameras

Here we combine the basic camera model (3.17) which gives the undistorted image coordinates of a mapped feature and the geometrical relations between the optical and navigation data (Fig. 2.5 and Sec. 2.5.2). Let $\mathbf{p}_n^m$ be the 3-D coordinates of the n-th tie-point expressed in the mapping frame. Considering the poses of b-frame associated with the j-image $\Gamma_{b,j}^m = [\mathbf{x}_b^m(j), \mathbf{R}_b^m(j)]$, through the camera boresight $\mathbf{R}_s^b$ and lever-arm $\mathbf{x}_{bs}^b$, this point is projected to the image plane of an ideal perspective camera with camera constant c and principal point (x_0, y_0) at image coordinates (x, y), such that

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \frac{1}{\mu} \begin{pmatrix} c & 0 & x_0 \\ 0 & c & y_0 \\ 0 & 0 & 1 \end{pmatrix} \left[\left(\mathbf{R}_b^m(j)\, \mathbf{R}_s^b \right)^T \left(\mathbf{p}_n^m - \mathbf{x}_b^m(j) \right) - \left(\mathbf{R}_s^b \right)^T \mathbf{x}_{bs}^b \right]. \tag{4.2}$$

The scale factor μ can be eliminated by rearranging the equation system (4.2) so that the image coordinates are separated on the left side and then dividing the first two relations. Then, the distorted image coordinates (x_d, y_d) can be determined, e.g., as in (4.1). Finally, with $\mathbf{l}_{n,i}$ being the image coordinate of the n-th 3-D point on the image, as reported by the tie-points detection algorithm, the image observation model reads

$$\mathbf{l}_{n,i} + \mathbf{v}_{n,i} = \begin{pmatrix} x_d \\ y_d \end{pmatrix} \tag{4.3}$$

where $\mathbf{v}_{n,i}$ is the correction vector.

Line cameras

The collinearity condition expressed for frame cameras (4.2) with the column (x) and row (y) image coordinates can be adapted to line cameras when omitting the y or row pixel coordinates:

$$\begin{pmatrix} x \\ 0 \\ 1 \end{pmatrix} = \frac{1}{\mu} \begin{pmatrix} c & 0 & x_0 \\ 0 & c & 0 \\ 0 & 0 & 1 \end{pmatrix} \left[\left(\mathbf{R}_b^m(j)\, \mathbf{R}_s^b \right)^T \left(\mathbf{p}_n^m - \mathbf{x}_b^m(j) \right) - \left(\mathbf{R}_s^b \right)^T \mathbf{x}_{bs}^b \right].$$

(4.4)

Nevertheless, most line cameras are multiple m-line cameras (with lines $k = 1, \ldots, m$). In such a case, with the use of projection matrix $\mathbf{K}$ (3.16), the undistorted collinearity model reads

$$\begin{pmatrix} x_k \\ 0 \\ 1 \end{pmatrix} = \frac{1}{\mu}\mathbf{K}\left[\left(\mathbf{R}_b^m\left(j\right)\mathbf{R}_p^b\mathbf{R}_{s,k}^p\right)^T\left(\mathbf{p}_n^m - \mathbf{x}_b^m\left(j\right)\right) - \left(\mathbf{R}_p^b\right)^T\mathbf{x}_p^b + \mathbf{x}_{bs,k}^p\right]. \tag{4.5}$$

where p is the common camera-platform reference frame, $\mathbf{x}_{bs,k}^p$ is a lever-arm from line k to p-frame origin, and $\mathbf{R}_{s,k}^p$ is the rotation from camera k-camera to platform frame. With $\mathbf{l}_{n,i}^k$ being the distorted image coordinate of the n-th 3-D point on the image, as reported by the tie-points detection algorithm for the line k, the image observation model reads

$$\mathbf{l}_{m,n,i} + \mathbf{v}_{m,n,i} = \mathbf{x}_{d,m}\,. \tag{4.6}$$

The relation between distorted x_k and undistorted image coordinates $x_{d,k}$ depends on the optical model described in Sec. 4.3, e.g., the first line of (4.1).

4.4.2 Ground control

If available, observations of the 3-D coordinates of n-th mapped feature are introduced as

$$\mathbf{l}_n + \mathbf{v}_n = \mathbf{p}_n^m\,. \tag{4.7}$$

4.4.3 Position

The position of the sensor is determined either by a GNSS receiver or by GNSS/INS integration. The first observation refers to the antenna phase center a, the later is usually the origin of the b-frame. Both are determined with respect to some global coordinate system, e.g., WGS-84, which can be transformed to m-frame. Referring to (2.5) and considering these positions $\mathbf{x}_{a/b}^m$ with respect to m together with the lever-arm $\mathbf{x}_s^{a/b}$ and attitude $\mathbf{R}_b^m$, the observation model for sensor position reads

$$\mathbf{l}_{p,j} + \mathbf{v}_{p,j} = \mathbf{x}_{a/b}^m\left(j\right) + \mathbf{R}_b^m\left(j\right)\mathbf{x}_s^{a/b} \tag{4.8}$$

When the chosen mapping frame is a projection and the sensor-fusion model is derived for a Cartesian frame, the observation of position should be corrected in height. As described in Legat (2006), the amount of such correction depends on the absolute terrain height, flying altitude above it and the value of projection-scale at perspective centre.

4.4.4 Velocity

Velocity observations are needed for sensors like Radar. They can be also useful for estimating a time-stamping offset between optical and navigation data (Rehak and Skaloud, 2017). GNSS or GNSS/INS provide velocity observation of the antenna a or body b either in e or l frame, respectively. Similarly to position, the velocity vector can be transformed to mapping frame. The velocity observation model is then

$$\mathbf{l}_v + \mathbf{v}_v = \mathbf{v}^m_{a/b}(j) + \mathbf{R}^m_b(j)\boldsymbol{\Omega}^b_{mb}(j)\,\mathbf{x}^{a/b}_s \tag{4.9}$$

where $\boldsymbol{\Omega}^b_{mb}(j)$ is the skew-symmetric matrix of angular velocity vector[5] between the m and b frames expressed in b-frame.

4.4.5 Attitude

Corrections to attitude observations $\mathbf{v}_R$ are expressed as non-commutative multiplication of rotation matrix $\mathbf{R}_v$. Hence, if attitude is observed externally as $\mathbf{R}^b_m$, the attitude observation equations reads

$$\mathbf{l}_R + \mathbf{v}_R \equiv \mathbf{R}^b_m\,\mathbf{R}_v = \mathbf{R}^b_s\,\mathbf{R}^s_m \tag{4.10}$$

with $\mathbf{R}^b_s$ being the boresight and $\mathbf{R}^s_m$ the mapping-to-sensor frame rotation. However, INS/GNSS processing usually delivers $\mathbf{R}^b_l$, where the orientation of the local-level frame l with respect to m-frame changes with the change of position. In such a case the attitude observations equations needs to be modified to

$$\mathbf{R}^b_l\,\mathbf{R}_v = \mathbf{R}^b_s\,\mathbf{R}^s_m\,\mathbf{R}^m_l\,, \tag{4.11}$$

where the definition of $\mathbf{R}^m_l$ depends on the arbitrary choice of the mapping frame. For instance, with mapping-frame defined as Cartesian system on a tangent-plane of WGS-84 ellipsoid at geographical coordinates (ϕ_0, λ_0)

$$\mathbf{R}^{m\equiv l_0}_l = \mathbf{R}^l_e(\phi_0, \lambda_0)\;\mathbf{R}^e_l(\phi_t, \lambda_t)\;, \tag{4.12}$$

where $\mathbf{R}^e_l = \left(\mathbf{R}^l_e\right)^T$ is defined by (2.2).

Conformal projections are often used when mapping in country-specific national coordinates. There, the convergence of meridian γ_{PC} at each perspective-center position (PC) needs to be accounted. Considering East-North-Up axis convention usually used n projections, the attitude observation for projection reads

$$\mathbf{R}^{m\equiv p}_l = \mathbf{R}^{ENU}_{NED}\,\mathbf{R}_3(\gamma_{PC})\;, \tag{4.13}$$

[5] $\omega = (\omega_1, \omega_2, \omega_3) = (\Omega_{3,2} = -\Omega_{2,3}\,, \Omega_{1,3} = -\Omega_{3,1}\,, \Omega_{2,1} = -\Omega_{1,2})$

where the matrix on the left-hand side involves exchanging the first and second axis and mirroring the third one and the second term is a standard rotation matrix about the third axis of the (modified) p-frame with the meridian convergence value γ_{PC} computed at the sensor perspective-center for the particular projection. When the mapping-frame is a conformal projection defined on a national reference ellipsoid, the observation equation for attitude may need to be further modified for the relative rotation between the ellipsoid employed for INS/GNSS integration and that of national datum (Skaloud and Legat, 2008).

4.4.6 Angular velocity and specific forces

As outlined in Sec. 4.1, some method of robotics's SLAM or dynamic networks, directly employs the inertial raw observations, i.e., the angular velocities ω and specific forces f. The rigorous form of these observation equations is rather long and is described in detail Cucci et al (2017). As schematically depicted in the bottom part of Fig. 4.5, ω and f constrain the unknown poses $\Gamma^m_{b,j}$ via differential relations that are approximated by first and second order finite differences. Also shown schematically in Fig. 4.5 are the connections between other previously mentioned observation models (represented by boxes) with the unknown parameters (represented by circles): p for GNSS positions, l^m_n for image observations, and 0 for so called zero-observations. The latter relates some parameters by known functional relationship without being associated with an actual sensor reading, e.g., interpolation between poses to image observation times 0_i, or time-correlated evolution of accelerometer biases 0_{bf}. Although usually applied, the evolution of gyroscope models $0_{b\omega}$ is not represented in Fig. 4.5 for the sake of clarity.

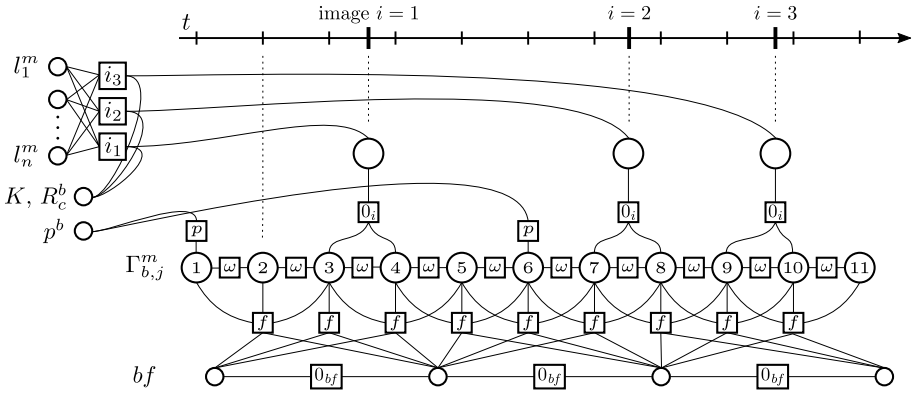


Figure 4.5: A simplified instance of a dynamic network formulated as a factor graph.

4.5 Estimation

The goal of estimation is in the optimal combination of all observations that leads to the most correct values of unknown parameters. Assuming that the correction or residual vectors $\mathbf{v}$ of all observations are randomly distributed, this is achieved by solving a non-linear weighted least-square problem. Gathering all the terms on one side of each observation equation will result in the following condition

$$g(\mathbf{l} + \mathbf{v}, \mathbf{x}) = 0 \tag{4.14}$$

where $\mathbf{l}$ represents the vector of given observation, $\mathbf{x}$ is the vector of unknown parameters and $\mathbf{v}$ is the vector of residuals of the observations that are assumed to be normally distributed, i.e. $\mathbf{v} \sim N(0, \mathbf{C}_{\ell\ell})$. Although the conditioning relation $g(\cdot)$ varies per sensors and observation, the general estimation methodology by least-squares principle stay the same.

- First, a linear model is obtained by linearizing the (non-linear) function $g(\cdot)$ according to the observations $\mathbf{l}$[6] and parameters $\mathbf{x}_i$, where the index $i = 0$ denotes its initial approximation

$$g(\mathbf{l}, \mathbf{x}_i) + \left(\frac{\partial g}{\partial \mathbf{l}}\right)_{|(\mathbf{l},\mathbf{x}_i)} v + \left(\frac{\partial g}{\partial \mathbf{x}}\right)_{|(\mathbf{l},\mathbf{x}_i)} \Delta\,\mathbf{x}_i = 0 \tag{4.15}$$

  or

$$-\mathbf{g}_i - \mathbf{B}_i\,\mathbf{v}_i + \mathbf{A}_i\,\Delta\,\mathbf{x}_i = 0 \tag{4.16}$$

  where $\mathbf{g}_i = -g(\mathbf{l},\, \mathbf{x}_i)$, $\mathbf{B}_i = -\left(\frac{\partial g}{\partial \mathbf{l}}\right)_{|(\mathbf{l},\mathbf{x}_i)}$ and $\mathbf{A}_i = \left(\frac{\partial g}{\partial \mathbf{x}}\right)_{|(\mathbf{l},\mathbf{x}_i)}$.

- Second, the non-linear model (4.14) is solved by iterating the solutions of the linear model (4.16) to convergence. The correction to the parameters $\Delta x_{\,i+1}$ are obtained by the Best Linear Unbiased Estimation (BLUE). As for the parameters, the BLUE estimation is the solution[7] of the so called normal equation (Förstner and Wrobel, 2004)

$$\Delta\mathbf{x}_i = \left(\mathbf{A}^T\left(\mathbf{B}\mathbf{C}_{vv}\mathbf{B}^T\right)^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^T\left(\mathbf{B}\,\mathbf{C}_{\ell\ell}\mathbf{B}^T\right)^{-1}\mathbf{g}_i\ , \tag{4.17}$$

  while the corrections to observations are obtained as:

$$\mathbf{v_i} = \mathbf{C}_{vv}\mathbf{B}^T\left(\mathbf{B}\mathbf{C}_{\ell\ell}\mathbf{B}^T\right)^{-1}(\mathbf{A}\mathbf{x} - \mathbf{g}_i)\ . \tag{4.18}$$

[6] in case of $\mathbf{v} = -\mathbf{l} + g(\cdot)$ this derivative is trivial

[7] In practice the matrices are rarely inverted explicitly for numerical and memory reasons.

After each iteration, the set of parameters is rectified by the estimated correction $\Delta\mathbf{x}_i$ as $\mathbf{x}_{i+1} = \mathbf{x}_i + \Delta\mathbf{x}_i$ and the observations are updated accordingly $\mathbf{l}_{i+1} = \mathbf{l}_i + \mathbf{v}_i$. The linearization of (4.15) is repeated with the updated set of parameters and observations ($x_{i=0}$ and $l_{i=0}$ denotes parameters and observation initial values, respectively). The iteration is stopped when the corrections to parameters $\Delta\mathbf{x}_i$ are not longer significant (i.e. $\Delta\mathbf{x}_i \sim 0$). After the convergence, the last iteration step is repeated with the original observations ($\mathbf{l}_{i=0}$). The respective covariances characterizing the accuracy of parameters and measurements are estimated in parallel at each step by relations presented in Förstner and Wrobel (2004).

The quality of the estimation may be judged according to the analyses of residuals and global a-posteriori estimation of the variance. The later is evaluated as $\widehat{\sigma}_0^2 = \left(\mathbf{v}^T\ \mathbf{C}_{\ell\ell}{}^{-1}\ \mathbf{v}\right)/(n-u)$, where n is the number of observations and u the number of parameters. Special situations may lead to some simplification of the general model (4.15) and its solution (4.17). Detailed information on this subject is presented in in Bjerhammar (1973) or in Förstner and Wrobel (2004).

The general formulation of the sensor fusion may be very large leading to hundreds of thousands unknowns (or even millions of unknowns for the dynamic networks[8]), but is inherently sparse and can be solved efficiently exploiting state-of-the-art least-squares solvers (Kummerle et al, 2011) based on very efficient sparse linear algebra routines (Davis, 2006).

4.6 Adopted approaches

4.6.1 Frame sensors

Advances in computer vision and digital-image processing enabled fully automated selection and measurement of corresponding points, which together with satellite positioning improved the productivity and accuracy of mapping. The stability of GNSS-assisted aero-triangulation remains dependent on the image texture, whose variation may cause problems in large-scale or oblique imagery, in forested areas or over snow-covered landscape. These problems can be somewhat mitigated with the the concurrent employment of integrated inertial navigation that allows also direct orientation. The latter concept found its place in fast mapping, applications of lower mapping accuracy, corridor mapping and terrestrial mobile mapping. Absolute orientation based solely on ground control points remain being used on small mapping projects, as those performed by small drones without RTK capacity.

GNSS are included for quality control or for calibration purposes on

[8]Reducing the number of unknowns is possible by pre-integrating certain number of IMU observations (Cucci and Skaloud, 2019).

Aerial obs.	Advantages	Disadvantages
None	⊕ Independent of airborne satellite signal ⊕ Simple processing chain ⊕ Independent of navigation quality ⊕ Independent of synchronisation errors	⊖ Impractically over large or steep areas ⊖ May lead to systematic deformations (func. of GCPs) ⊖ IO correlated to EO ⊖ Large overlaps required
Position	⊕ Absorbs IO instability ⊕ Consistent determination of all parameters ⊕ Potential for radiometric adjustment ⊕ Self-calibrating and possibly no GCPs	⊖ Week geometry at block ends ⊖ Not ideal for corridors ⊖ Larger side-overlap required ⊖ Textureless (e.g. costal) mapping is difficult ⊖ Problematic transfer of points in oblique-imagery
Full	⊕ Suitable for corridors & multi-sensor systems ⊕ ≈20% side-overlap OK ⊕ Automation, no GCPs	⊖ Lower redundancy in corridors ⊖ Attitude accuracy dependent

Table 4.2: Comparison between main orientation approaches.

larger missions benefiting navigation technology. Indeed, when factors such as accuracy and reliability are important, the method of integrated sensor orientation remains the most sophisticated alternative for frame-camera orientation (Tab. 4.2). In this method, the first approximation of exterior orientation is provided by the navigation technology, which is present on all modern large-scale digital cameras and auto-piloted drones[9]. Knowledge of the initial EO limits the search space for homologous points and thus improves their transfer between images on challenging texture. In this regard, external knowledge of attitude parameters is more important for oblique photography or situations with corresponding image texture then for vertical configurations and where sufficient image texture is available. In the next step, the optimization is run first to eliminate outlier observations, and later to provide the final solution to the orientation problem. Ground control points are included for quality control or for calibration purposes. Similarly to position-assisted AT, the use of full aerial control result in lower

[9]for the purpose of navigation, guidance and control capacity on automated missions

correlation between EO/IO parameters. Tab. 4.3 indicates the common orientation approaches across different platforms. Generally, it is acknowledged that object-space accuracy is 2–4 times better when using integrated rather than direct sensor orientation approach.

		Satellite	**Aircraft**	**Vehicle**
On-board sensors	GNSS	++	+++	+++
	IMU	++	++	+++
	Star Tracker	+++	+	–
External measurements	GCP	+++	++	+
Orientation approach	Direct	+	++	+++
	AAT/(GNSS)	++	+++	+
	Integrated	+	++	+

Table 4.3: Indicative frequency of sensor deployment and used orientation method for frame-cameras across different platforms: (–) = never, (+) = rare, (++) = sometimes, (+++) = common.

4.6.2 Line sensors

Theoretically, orientation of line sensor data can also be performed indirectly, in a similar manner to frame imagery (Hoffman et al, 1982). However, this approach is rarely used in practice, because the computational effort is large and the resulting mapping accuracy is lower than with the support of attitude and position sensors (Cramer, 2006). Also, to ensure sufficient overlap between successive exposures, the line camera head needs to be placed on a stabilized mount. Such stabilization can be more precise when based on real-time GNSS/INS trajectory, which is also the case for modern line cameras.

The common approach to line camera orientation is depicted in Fig. 4.6. The on-board GNSS/INS measurements are recorded for post-processing (PPK) and integration. At the same time, a real-time navigation solution based on point-positioning GNSS/INS integration is used to steer the camera-platform stabilization. The captured images are stored and rectified in post-processing using the best available calibration parameters and the improved EO parameters from post-processed trajectory. The distortion in the original imagery, caused by motion of the sensor, is removed by this rectification, and the resulting scenes can be viewed stereoscopically. The automated matching process is performed, but the tie-points coordinates are referred back to the original imagery. The orientation parameters are updated by the block adjustment using image measurements and orientation parameters. Possibly, ground control points (GCPs) may also be included for calibration, improved accuracy, or for control purposes. The block adjustment provides final orientation parameters that are applied either to

the prerectified images for DEM (digital elevation model) generation or also directly to the the raw images for the (best possible) orthophoto production.

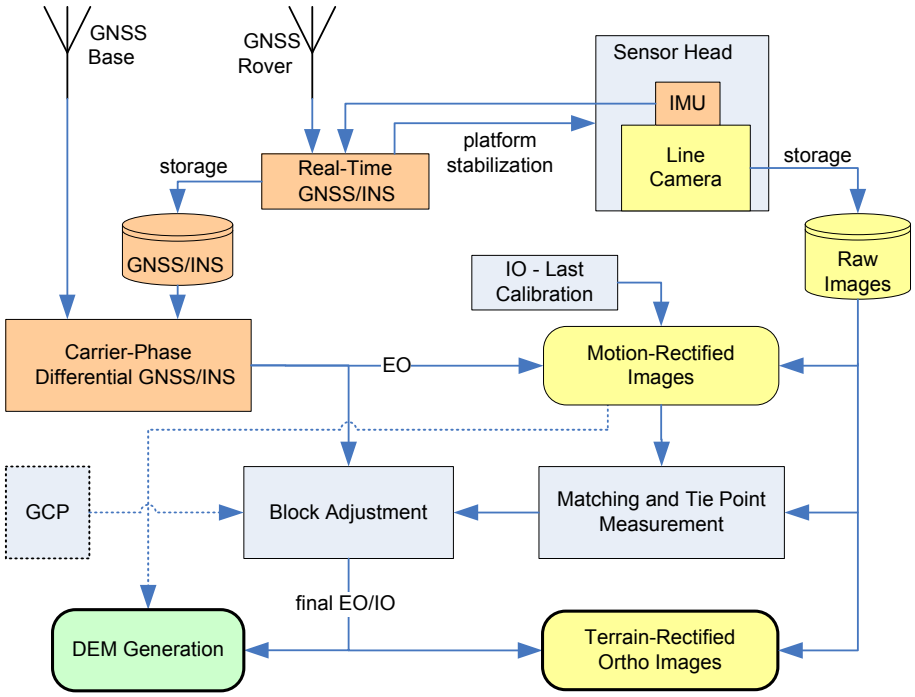


Figure 4.6: Processing chain for line-camera data.

Investigations with modern line cameras have revealed that, when a real-time GNSS/INS solution is used to support the matching process, the number of matched points is approximately 25% less compared with uses post-processed trajectory (Cramer, 2006). Also, the accuracy of object points was 2–3 times worse when using the EO parameters based on real-time trajectory, which is not acceptable for applications with highest accuracy demands (i.e., ≤ 0.1 m in object space). It was also observed that the mapping accuracy is 2–4 times worse when based on direct orientation as compared to the integrated sensor orientation.

4.6.3 Calibration

As soon as sensors such as GNSS were added to derive the camera perspective center coordinates in order to directly measure the exterior orientation elements with high absolute accuracy, systematic differences between perspective center coordinates derived from a bundle adjustment without "assistance" and the directly measured coordinates are likely to occur. Those differences cannot always be attributed to errors in the trajectory computations, especially when GNSS trajectory solutions be delivered with high

accuracy. Thus, such differences also might be caused by changes in the camera geometry: For example, since airborne images are mostly taken in nadir direction, incorrect assumptions of the focal length will shift the adjusted perspective center coordinates along the vertical axis. This immediately causes offsets between the directly measured perspective center coordinates and the coordinates obtained from the photogrammetric bundle adjustment. On the other hand, the availability of direct exterior orientation measurements of sufficiently high accuracy now offers the possibility to completely calibrate the camera geometry based on in situ approaches even for airborne sensors. However, a flat field such as the Earths surface combined with parallel viewing directions does not allow for determination of the full camera geometry unless the test field is of special design (e.g., it has significant height variations). Such requirements are necessary to suppress the high correlations between unknown exterior orientation elements and certain sensor parameters (i.e., sensor interior orientation). The additional sensor parameters are estimated in extended aerial triangulation. Additional parameter models which directly relate to physical changes in the sensor geometry are well established in photogrammetric imaging Förstner and Wrobel (2004) but have rarely been used for airborne camera calibrations in the past. Due to the previously described correlations, mostly mathematical polynomials were preferred to overcome remaining systematic effects in airborne imagery. Such parameter sets have been proposed by Ebner (1978) and others. These additional parameters are not correlated with the exterior orientation elements and can thus be used in standard aerial triangulation, but they do not refer to changes in the camera geometry. Another aspect is the need to calibrate sensor systems instead of single system components only. This is also referred to as system versus component calibration and becomes obvious if the design of today's digital imaging sensors is considered. They typically consist of several components

- The imaging sensor itself, which may contain several optical lens systems.
- Additional sensors for direct measurement of the sensor trajectory during data capture, which are almost standard for new digital sensors.

In contrast to film-based cameras, where calibration mainly considered the lens component only, the overall calibration of such, more complex systems cannot be done from laboratory calibrations exclusively. The relative orientations between the optical sensor and inertial measurement unit can only be derived from in situ approaches. This method is also convenient to derive the relative positions between GNSS-antenna, inertial and camera perspective center.

4.6.4 Laser scanners

The process of kinematic laser scanning relies on direct sensor orientation. Nevertheless, the principle of integrated sensor orientation can be introduced either for system calibration (Skaloud and Lichti, 2006), for the mitigation of residual systematic errors in trajectory determination (Filin and Vosselman, 2004) or for both (Kager, 2004; Friess, 2006; Glira et al, 2019). Such an adjustment process also servers as an internal control of the laser-scanning mission. The initial development of block adjustment in kinematic laser-scanning used the concept of tie-points[10]. Contrary to cameras, this principle is not very suitable as the correspondence between laser points is only approximate. The modern approaches therefore rely on conditioning surface patches or other geometrical primitives that overlap between different passes (Fig. 4.7). The success of this approach depends on the number of patches and their form, size and spatial variation. Generally, this approach works better on patches, whose form is known a-priori. This is the case for planar surfaces on buildings or other man-made structures. The parameters of the planes are estimated together with the calibration parameters that may include also biases in the trajectory (Glira et al, 2019). Such trajectory bias modelling is approximate and can be avoided when the estimation includes the inertial raw readings as observations (Cucci et al, 2017) (c.f., Sec. 4.4.6). When considering the simpler formulation with the platform position and orientation provided by a GNSS/INS, together with the range and encoder angle values measured by the laser-scanner, there are eight observations per laser return. Using (2.6), the observation equation for a laser target in the e-frame $\mathbf{x}_i^e$ lying on a plane $\mathbf{s}_j$ is given by

$$\left\langle \mathbf{s}_j \, , \, \begin{pmatrix} \mathbf{x}_i^e \\ 1 \end{pmatrix} \right\rangle = 0 \, , \tag{4.19}$$

where, the plane parameters are given by

$$\mathbf{s}_j = (s_1 \;\; s_2 \;\; s_3 \;\; s_4)^T \, . \tag{4.20}$$

with s_1, s_2, s_3 being the direction cosines of the plane's normal vector and s_4 the negative orthogonal distance between the plane and the coordinate system origin. Note that the direction cosines must satisfy the unit length constraint $\|\mathbf{s}_j\| = 1$. The details how such constrain is added to the adjustment model are described in Skaloud and Lichti (2006). The principle can be extended for natural surfaces (Filin, 2003), however, this approach has certain limits. First, most of the naturally flat surfaces are horizontal which makes their contribution less significant. Second, perfectly flat surfaces are less common in nature and their identification remain problematic

[10] based on return-intensity values

(Kerstling et al, 2012). The future methods of integrated sensor orientation with laser data will most likely start using general surface models with somewhat tighter feedback to the trajectory determination as it is the case of robotics SLAM (Strasdat et al, 2010). However, as terrestrial robots usually employ multi-beam 3D scanners, there is only one set of pose parameters per one instance of beam-array activation. This configuration is somewhat similar to line-cameras, and geometricaly stronger than the more usual case of airborne scanning, where each single laser pulse has a unique set of pose.

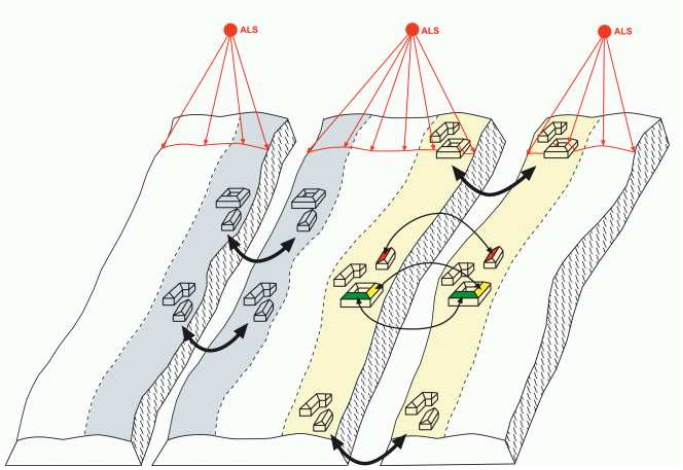


Figure 4.7: Principle of surface-patch conditioning in airborne laser scanning, after Kager (2004).

Chapter 5
Mapping Products

5.1 Surface

In this section remote sensing is used in a general sense and thus incorporates photogrammetry. The products of remote sensing may be grouped by their production process in geometric and radiometric products. Here we briefly introduce two typical photogrammetric products, one in 3-D and the other in 2-D: surface models and the orthophoto.

5.1.1 Representation

The term digital elevation model (DEM) encompasses surface representation without specifying its nature. On the other hand, digital terrain model (DTM) is a discrete description of the physical surface (terrain), while the digital surface model (DSM) considers the terrain with all surface features including buildings and vegetation. The captured information about the terrain height by means of optical sensors (Sec. 1) is usually heterogeneous and unorganized. Therefore, it needs to be restructured into a form that is both comprehensive and usable for further exploitation like interpretation, visualization, manipulation, etc.

Regardless of its form, a surface model will always remain an approximation of the reality with limited resolution. However, the choice on its representation is important as it dictates the requirements on data storage, possibility to portray sharp changes in the topography or the efficiency in model manipulation and analysis. Between the number of possibilities on terrain representation, the grid, triangles and contours are the most common and therefore will be discussed in more detail (Fig. 5.1).

Elevation grid

An elevation grid is the most straightforward representation of the terrain (Fig. 5.1C). It is characterized by regular, lattice organization of equally spaced points in the horizontal (x, y) projection. Each point of such mesh contains one height (z) value for its location that is together with x, y coordinates referenced to common origin. The spacing between points is predefined and thus implies the resolution of the model. Such organization is similar to an image and due to such resemblance this representation is referred to

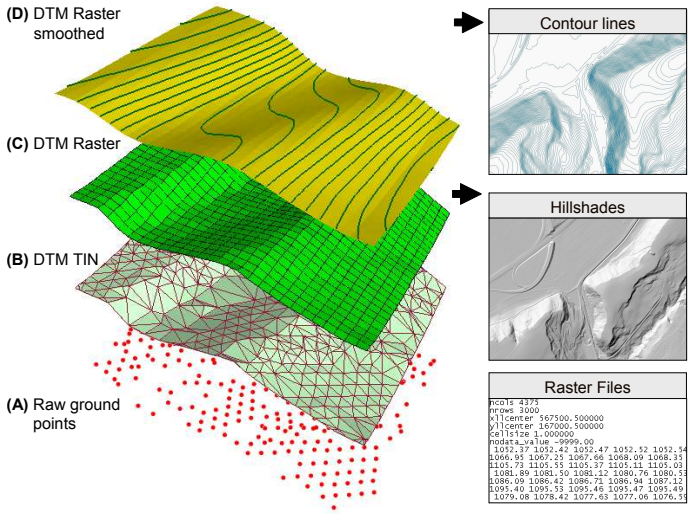


Figure 5.1: Typical surface modeling based on irregularly distributed points (A). (B) Generation of a triangular irregular network (TIN).(C) Interpolation of TIN to regularly spaced grid. (D) Smoothing of grid and derivation of contour lines, after Schaer (2009).

as raster. In this structure, only z values need to be stored as the x, y coordinates are derived from corresponding indices and cell-spacing.

The grid structure is convenient for its simplicity in organization that is also practical for further data manipulation. On the other hand it is less suitable for modelling of steep landscapes where the resolution with respect to slope normally decreases progressively (Tab. 5.1). It is also not suitable for modelling complex shapes in three dimensions as overhangs, because the storage of several z values per grid-cell is not possible. The grid arrangement is also sub-optimal in capturing characteristic landscape features like highest points or break-lines which may not coincide with a grid cell. To describe finer terrain features by this method the cell-size needs to be reduced. However, this increases storage requirements without providing additional information in areas where coarser cell-size is adequate. Such inconvenience could be circumvented by allowing the cell-size to be adaptive (e.g. quadtree storage) or by applying image-like compression (El-Sheimy et al, 2005).

TIN

Compared to a grid, the triangulated irregular network (TIN) structure (Fig. 5.1B) is considerably better adaptive to local terrain variations. It constitutes a set of nodes (points) that are connected by lines to form triangles. The surface within each triangle is represented by a plane facet. The individual facets fits in a mosaic that yields a surface. Such modelling is appropriate to areas with sudden changes in slopes, where the edges of triangles can be aligned with discontinuities in the landscape (e.g. ridges, bottoms of gullies).

The storage requirements of irregular-spaced points and its associated TIN structure are quite large, as all three coordinates per point need to be stored separately. A solution that overcomes this problem is the beforementioned generation of uniformly-spaced elevation grids, where the x, y coordinates are described by an array of indexes and only the z-value is stored.

Contours

Terrain representation by contours (Fig. 5.1D) was the most common way of surface modelling before the digital era. It is also the most frequent means of coding the vertical dimension into topographical maps. Contours are lines of constant elevation (isolines) that are usually projected on a 2-D surface. In the past contours were generated manually from oriented photographs on stereo-plotters. Although laborious, this process was accurate when carried out by a skilled operator who, at the same time, made judicious generalization of reality. Although the DTM could be derived from contours by interpolation, this practice is left to cases when a cartographic source (i.e., map) is the primary input for its generation. In modern mapping the contours are produced automatically from a grid, TIN or irregularly distributed elevation points (El-Sheimy et al, 2005).

A comparison between different forms of terrain representation is given in Tab. 5.1.

5.1.2 Reconstruction

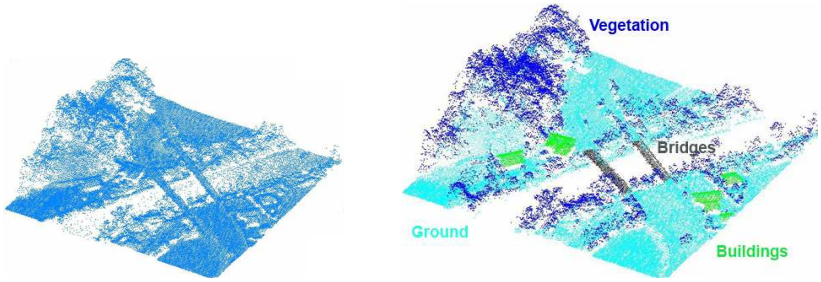
DEMs of coarse resolution covering all continents are mostly produced by satellite missions (InSAR). The acquisition of DEM of finer resolution at the scale of large countries is most effectively performed by airborne SAR, less effectively but more accurately by image processing using the principles presented in Sec. 4. The altimetric models of highest precision and resolution usually come from airborne laser scanning.

	Grid	**TIN**	**Contours**
Structure	⊕ Simple	⊖ Complex	⊖ Complex
Storage	⊕ More compact	⊖ Larger	⊖ Large
Exchange	⊕ Excellent	⊖ 2.5D possible	⊖ Difficult
Applicability	⊖ Limited 2.5D	⊕ 3D possible	⊖ Limited 2.5D
Adaptability	⊖ With quad trees	⊕ Adaptable	⊕ Adaptable
Modeling	⊖ With sampling rate	⊕ Excellent	⊖ Modest
Discontinuity	⊖ Limited	⊕ Good	⊖ Limited
Operations	⊕ Fast and robust	⊖ More complex	⊖ Not practical
Usage in maps	⊕ As hillshade	⊖ Not practical	⊕ Excellent

Table 5.1: A comparison between different forms of DEM representation; ⊕ and ⊖ denote advantages and disadvantages, respectively.

Classification

As laser scanning is a non-selective mapping method, the acquired point-cloud includes all kinds of objects (e.g., vegetation, buildings, wires) apart from the terrain itself. Hence, prior to the derivation of elevation models the point-cloud needs to separated into categories of objects as depicted in Fig. 5.2. This process is called classification and is highly but not entirely automated. When the point data are obtained by insertion from oriented images (Sec. 4), the classification could be performed together with image matching.



(a) Raw laser point-cloud,

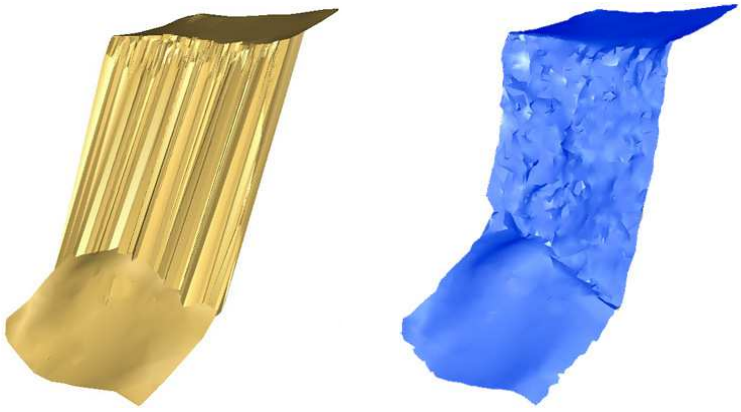
(b) Point-cloud classified to ground, bridge, vegetation and building points.

Figure 5.2: Point cloud classification, after Schaer (2009).

Triangulation

As mentioned previously, triangulation creates a polygonal or triangular mesh from a set of unorganized points, where the facets of polygons are the discrete representations of the surface (Fig. 5.1-B). Triangulation can be

performed in 2-D or in 3-D, according to the geometry of input data. Large, country-like elevation models are created usually in 2.5-D, which means that the triangulation is performed in 2-D and the z-value gets attached to each node using a unique elevation function $z = f(x, y)$. Such models are not ideal to represent steep terrain (Fig. 5.3-left) or complex man-made structures. As can be seen from the right part of the Fig. 5.3, performing 3-D triangulation is more suitable for this purpose, however, its evaluation is very complex in large data sets. Also, data exchange in the GIS community is not necessary standardized for 3-D TIN structures.



(a) 2.5-D triangulation. (b) 3-D triangulation.

Figure 5.3: Isometric view of a vertical rock face obtained by airborne laser scanning and triangulation; after Schaer (2009).

Grid generation

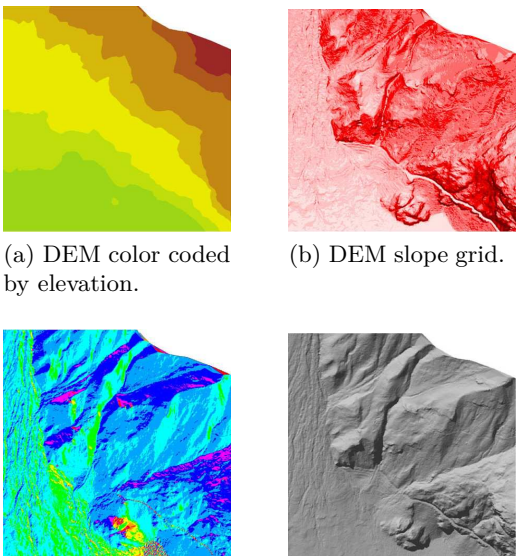
An evenly-spaced DTM grid can be obtained by interpolation from the TIN facets. This approach is, however, practical only when the TIN model already exists. The grid can also be derived directly from the point data by various techniques of interpolation. The popular interpolations methods used for this purpose are trend surface analysis, Fourrier analysis or Kriging. These approaches have a global character and various level of smoothing that is either predetermined or estimated from the data itself. On the other hand, methods of local character are more appropriate when the terrain varies abruptly as they are based on the elevation information from the nearest points. The most frequent methods of such type are spline or cubic interpolations and inverse distance weighting (El-Sheimy et al, 2005).

In places where the ground point sampling is low, like in forested areas or within dense urban zones, it is preferable to apply triangulation prior to grid generation, because the resulting DEM is less affected by the lack of data. The maximal size of data gaps to be closed by triangulation can be

limited by specifying the largest length of the facet edge.

5.1.3 Analysis

The grid representation of the DEM allows applying image-like operations that are useful for highlighting different aspect of the surface. The most common are filters that perform terrain smoothing and gradient operations for visualizing steepness and orientation as shown in Fig. 5.4A-C. Fig. 5.4D depicts so called hillshade raster that improves visualization of the terrain by a chosen source (in position and angle) of illumination.



(a) DEM color coded by elevation.

(b) DEM slope grid.

(c) DEM color coded by aspect.

(d) DEM hillshade grid.

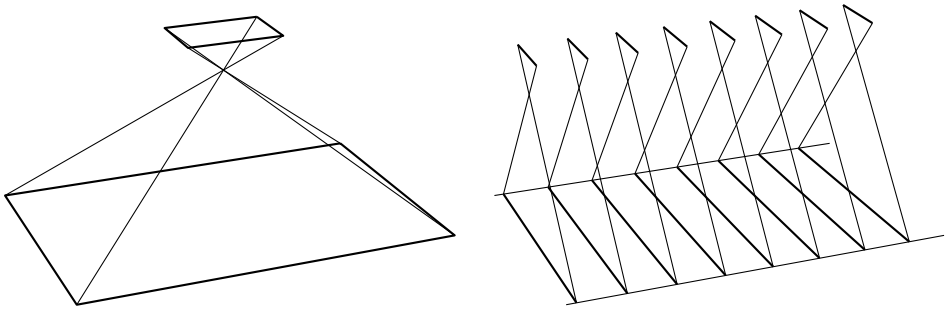
Figure 5.4: Analysis of raster height elevation model; after Schaer (2009).

5.2 Orthophoto

5.2.1 Orthogonal perspective

Orthophoto is a technical term reserved for an image that shows objects on a reference surface using an orthogonal perspective. The reference surface is a DEM that consists of points with three coordinates each (x, y, z) and that defines the Earth's surface. The orthogonal perspective means vertical view on the ground above each pixel, which is typically used for maps. However, any image taken by a camera does not have this perspective, because the whole scene is photographed from one point, thus leading to central perspective. Frame cameras have a central perspective in x and y (Fig. 5.5-a). Line

cameras operating on satellite platforms and in some cases also on aerial platforms produce images that have central perspective along the line photographed at one moment but orthogonal perspective along the flight track (Fig. 5.5-b).



(a) Frame camera with central perspective in both dimensions.

(b) Line camera with central perspective across track and orthogonal perspective along track.

Figure 5.5: Central perspective.

An aerial image is distorted primarily for two reasons

- The camera cannot be kept exactly horizontal when the photo is taken. Therefore, the roll and pitch components of attitude are not exactly zero, and consequently the image suffers from perspective distortion.
- Only in some exceptional cases is the Earth's surface flat. The central perspective of a camera causes height parallaxes that displace objects on higher ground towards the image border and objects on lower ground towards the image center.

Both effects are eliminated during orthophoto computation. In addition, the pixel size is set to a defined ground sample distance (GSD). This simplifies joint processing with vector data in later applications.

Other image errors such as lens distortion, atmospheric refraction, and Earth curvature are not eliminated during orthophoto computation, as their influence is usually considered in a basic part of the photogrammetric image-processing process that includes sensor calibration. Despite that, the geometry of the camera plays a prominent role in the computation of an orthophoto.

5.2.2 Rectification methods

The orthographic projection is obtained from the central perspective through analytical process called rectification. The complexity of this operation depends on the scale of imagery and the required degree of exactness. An overview of the different approaches is provided in Tab. 5.2.

Rectification Method	**Elevation Model**	**Application**
Perspective transformation	No model	Analytical plotter (approximation)
Polynomials		Satellite imagery
Standard orthophoto	DTM	Airborne imagery
True orthophoto	DTM + buildings	Airborne imagery (special application)

Table 5.2: Overview of the rectification methods.

The perspective transformation provides only an approximate solution to rectification and has been employed in the past when airborne photogrammetry made use of analog rectification instruments. The polynomials provide somewhat better 2-dimensional relation between the image and the ground. Such approximation is usually sufficient for the rectification of satellite images. The transformation coefficients are commonly obtained through the identification of ground control-points that are distributed across the image. This way, the images are simultaneously oriented and rectified.

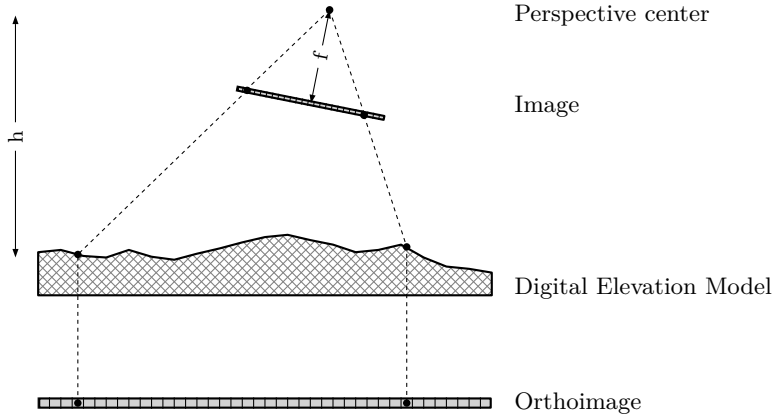


Figure 5.6: The creation of an orthoimage with the use of Digital Elevation Models (DEM).

By far most orthophotos are produced by making use of DEM (Sec. 5.1.1), especially DTMs which exclude vegetation and man-made structures. Consequently, the imaged buildings are rectified only at the base and not above the ground. Apart from the existence of the DTM, the prerequisite for correct rectification is the knowledge of image orientation parameters (interior and exterior) that shall be transformed to the same datum as the employed DTM. A common procedure for orthophoto creation using DEM is the following:

1. An empty orthophoto is created at the start. Such orthophoto can be regarded as a grid or matrix with cells of predefined (pixel) size. Hence, knowing the coordinates (x and y) of the image-corner point, the geographical position of each pixel is uniquely defined by its row and column.

2. The geographical height (z coordinate) of each pixel is identified through DEM as a function its x and y position (e.g. by interpolation).

3. A vector is formed between the image origin and the x-y-z coordinates of an empty pixel in the orthophoto (Fig. 5.6). This vector is intersected with the image through the collinearity condition to define its x' and y' photo-coordinates.

4. As the resulting x' and y' photo-coordinates do not necessary correspond to the center of a pixel, its RGB-color (or gray) value on the orthophoto is found via interpolation with the neighbouring pixels on the photograph. The interpolation can be performed also across several images and these values can be further averaged to stabilize the resulting orthophoto radiometrically.

Today, a *true orthophoto* is computed based on a dense point-cloud that represents a surface-model, i.e. mainly ground, buildings, and vegetation. Out of that point-cloud, all those points are deleted which are not the highest at any given position. The remaining points form the *true orthophoto* (Fig. 5.7).

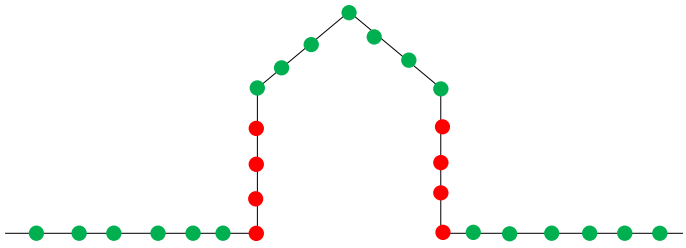


Figure 5.7: Creation of a true orthophoto based on a dense point cloud (3-D). Green points: Points of the true orthophoto; red points: Points not used for the true orthophoto because they are not highest at their 2-D-position.

Bibliography

Andersson R, Bilger H, Stedman G (1994) Sagnac effect: A century of earth-rotated interferometers. Am J of Phys (62):975–985

Betz J (2016) Engineering satellite-based navigation and timing. Wiley and IEEE Press

Bjerhammar A (1973) Theory of errors and generalized matrix inverses. Elsevier, Amsterdam

Burman H (2000) Calibration and orientation of airborne image and laser scanner data using gps and ins. PhD thesis, Royal Institute of Technology

Carabajal C, Harding D, Luthcke, SB Fong W, Rowton S, Frawley J (1999) Processing of shuttle laser altimeter range and return pulse energy data in support of lsa-02. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 32, Part 4A:269–277

Colomina I, Blazquez M (2004) A unified approach to static and dynamic modeling in photogrammetry and remote sensing. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 35-B1:178–183

Cosandier D (1999) Generating a digital elevation model and orthomosaic from pushbroom imagery. Ucge report no. 20133, The University of Calgary

Cramer M (2006) The ads40 vaihingen/enz geometric performance test. ISPRS Journal of Photogrammetry & Remote Sensing 60:363–374

Cramer M, Stallmann D (2002) System calibration for direct georeferencing. In: Photogrammetric Computer Vision, ISPRS Commission III Symposium, Graz, Austria, p 6

Cucci D, Skaloud J (2019) On inertial measurements in dynamics networks. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences IV-2/W5:549–557

Cucci DA, Rehak M, Skaloud J (2017) Bundle adjustment with raw inertial observations in uav applications. ISPRS Journal of Photogrammetry Engineering and Remote Sensing 130:1–12

Davis TA (2006) Direct methods for sparse linear systems. SIAM

Dorstel C, Jacobsen K, Stallmann D (2003) Dmc-photogrammetric accuracy - calibration aspects and generation of synthetic dmc images. In: Proceedings on Optical 3-D Measurements Techniques, Zurich, vol VI, pp 74–82

Ebner H (1978) Self calibrating block adjustment. In: XIIIth Congress of the International Society of Photogrammetry, Com. V, Stockholm, Sweden

Ehlers M, Klonus S, Astrand P, Rosso P (2010) Multi-sensor image fusion for pansharpening in remote sensing. Int J Image Data Fusion 1:25–45

El-Sheimy N, Valeo C, Habib A (2005) Digital terrain modeling. Artech House

Filin S (2003) Recovery of systematic biases in laser altimetry data using natural surfaces. Photogrammetric Engineering and Remote Sensing 69(11):1235–1242

Filin S, Vosselman G (2004) Adjustment of airborne laser altimetry strips. In: ISPRS Congress, International Archives of Photogrammetry and Remote Sensing, Istanbul, Turkey, vol 34, pp 285–289

Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun 24:381–395

Förstner W, Wrobel B (2004) Mathematical concepts in photogrammetry. In: McGlone J, Mikhail E (eds) Manual of Photogrammetry, 5th edn, American Society of Photogrammetry and Remote Sensing, Bethesda, MA., pp 15–180

Freyer J, Brown D (1986) Lens distrotion for close-range photogrammetry. Photogrammtric Engineering and Remote Sensing 52:51–58

Friess P (2006) Toward a rigorous methodology for airborne laser mapping. In: EuroCOW, on CDROM, Castelldefels, Spain, p 7

Geiger A, Kahle HG, Limpach P (2009) Airborne laser profiling. ETH Research database 6120, Swiss Federal Institute of Technology Zurich (ETHZ)

Glennie C, Lichti D (2010) Static calibration and analysis of the velodyne hdl-64e s2 for high accuracy mobile scanning. Remote Sensing 2:1610–1624

Glennie C, Kusari A, Facchin A (2016) Calibration and stability analysis of the vlp-16 laser scanner. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol XL-3/W4, pp 55–60, DOI 10.5194/isprsarchives-XL-3-W4-55-2016

Glira F, Pfeifer N, Mandelburger G (2019) Hybrid orientation of airborne lidard point clouds and aerial images. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences IV-2/W5:567–574

Gonzalez R, Woods R (1992) Digital image processing. Addison-Wesley

Greenspan R (1995) Inertial navigation technology from 1970 to 1995. Journal of The Institute of Navigation 42:165–186

Gruen A (1982) The accuracy potential of the modern bundle block adjustment in aerial photogrammetry. Photogrammetric Engineering and Remote Sensing 48:45–54

Haala N (2014) Dense image matching final report. Tech. rep., EuroSDR Official Publication NÂ°64, pp. 115-145

Haralick RM, Shapiro LG (1992) Computer and robot visions. Addison-Wesley, Reading

Harris C, Stephens M (1988) A combined corner and edge detector. In: Proc. 4th Alvey Vision Conference, pp 147–151

Hartley RI (2012) In defence of the 8-point algorithm. IEEE Trans Pattern Analysis 19(6):580–593

Hirschmüller H (2008) Stereo processing by semi-global matching and mutual information. IEEE PAMI 30(2):328–341

Hoffman O, Nave P, Ebner H (1982) Dps a digital photogrammetric system for producing digital elevation models and orthophotos by means of linear array scanner imagery. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 24(B3):216–227

Hug C (1994) The scanning laser altitude and reflectance sensor - an instrument for efficient 3d terrain survey. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 30:100–107

Jazayeri I, Fraser C (2010) Interest operators for feature-based matching in close range photogrammetry. Photogramm Rec 25(129):24–41

Jekeli C (2001) Inertial navigation systems with geodetic aapplications. Walter de Gruyter, Berlin

Kager H (2004) Discrepancies between overlapping laser scanning strips - simultaneous fitting of aerial laser scanner strips. In: Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol 35-B1, pp 555–560

Kerstling A, Habib A, Bang KI, Skaloud J (2012) Automated approach for rigorous light detection and ranging system calibration without preprocessing and strict terrain coverage requirements. Optical Engineering 51(7):076,201, DOI doi:10.1117/1.OE.51.7.076201, URL http://dx.doi.org/10.1117/1.OE.51.7.076201

Khaghani M, Skaloud J (2016) Application of vehicle dynamic modeling in uavs for precise determination of exterior orientation. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 23 ISPRS Congress, vol XLI-B3, pp 827–831, DOI doi:10.5194/isprsarchives-XLI-B3-827-2016

Khaghani M, Skaloud J (2018) Assesment of vdm-based autonomous navigation of a uav under operational conditions. Robotics and Autonomous Systems minor review pending

Kraus K (2007) Photogrammetry - Geometry from Images and Laser Scans. de Gruyter

Kruck E (2001) Combined imu and sensor calibration with bingo-f. In: Integrated Sensor Orientation, Proc. of the OEEPE Workshop ", CD-ROM, Hannover, pp 84–108

Kummerle R, Grisetti G, Strasdat H, Konolige K, Burgard W (2011) g^2o: A general framework for graph optimization. In: Robotics and Automation (ICRA), 2011 IEEE International Conference on, IEEE, pp 3607–3613

Ladstadter R, Tschemmernegg H, Gruber M (2010) Calibrating the ultracam aerial camera systems, an update. In: Proceedings International Calibration and Orientation Workshop, EuroCOW, p 8

Legat K (2006) Approximate direct georeferencing in national coordinates. ISPRS Journal of Photogrammetry & Remote Sensing 60:239–255

Lohr U, Beraldin A, Blais F (2010) Airborne and terrestrial laser scanning, Whittles Publishing, chap Laser scanning technology

Longuet-Higgins (1981) A computer algorithm for reconstructing a scene from two projections. Nature 293:133–135

Lowe DG (2004) Distinctive image features from scale invariant key points. Int J Comput Vis 60(2):91–110

Mikolajczyk K, Schmid C (2005) A performance evaluation of local descriptors. IEEE PAMI 27(10)

Morin K (2002) Calibration of airborne laser scanners. M.sc., UCGE Report No. 20179, The University of Calgary

Morin K, El-Sheimy N (2002) Post-mission adjustment of airborne laser scanning data. In: FIG XXII International Congress, Washington DC, USA, vol CD-ROM, p 12

Nistér D (2004) An efficient solution to the five-point relative pose problem. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(6):756–770

Petri G (2000) Deja vu - the configurations of the new airborne digital imagers are all rooted in the distant past! GeoInformatics 3(5):48–51

Petri G, Walker S (2007) Airborne digital imaging technology: a new overview. Photogrammetric Record 22:203–225

Rehak M, Skaloud J (2015) Fixed-wing micro aerial vehicle for accurate corridor mapping. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-1/W4:23–31, DOI 10.5194/isprsannals-II-1-W1-23-2015, URL http://dx.doi.org/10.5194/isprsannals-II-1-W1-23-2015

Rehak M, Skaloud J (2017) Time synchronization of consumer cameras on micro aerial vehicles. ISPRS Journal of Photogrammetry & Remote Sensing 123(1):114–123

Satirapod C, Homniam P (2006) Gps precise point positioning software for ground control point establishment in remote sensing applications. Journal of Surveying Engineering (ASCE) 132(1):11–14

Schaer P (2009) In-flight quality assessment and data processing for airborne laser scanning. PhD thesis, The Swiss Federal Institute of Technology Lausanne (EPFL)

Scherzinger B (2006) Precise robust positioning with inertially aided rtk. NAVIGATION, Journal of The Institute of Navigation 53(2):73–83

Schmid C, Mohr R, Bauckhage C (2000) Evaluation of interest point detectors. Int J Comput Vis 37(2):151–172

Skaloud J, Legat K (2008) Theory and reality of direct georeferencing in national coordinates. ISPRS Journal of Photogrammetry & Remote Sensing 63:272–282

Skaloud J, Lichti D (2006) Rigorous approach to bore-sight self calibration in airborne laser scanning. ISPRS Journal of Photogrammetry and Remote Sensing 61:47–59

Skaloud J, Schaer P (2003) Towards a more rigorous boresight calibration. In: ISPRS International Workshop on Theory Technology and Realities of Inertial/GPS/Sensor Orientation, Castelldefels, Spain

Spikes V, Csatho B, Whillans I (1999) Airborne laser profiling of antartic ice stream for change detection. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 32, Part 3-W14:7

Strasdat H, Montiel J, Davison A (2010) Real-time monocular slam: why filter? In: IEEE International Conference on Robotics and Automation (ICRA), pp 2657–2664

Szeliski R (2010) Computer vision: algorithms and applications. Springer

Tempelmann U, Hinsken L (2005) Triangulation of ads40 pushbroom image blocks - not much different from classical frame blocks?

Titterton D, Weston J (1997) Strapdown inertial navigation technology. Part of IEE radar, sonar, navigation and avionics series, Stevenage, U.K., all the details in strapdown inertial navigation, attention: some mistakes in equastions

W Förstner EG (1987) A fast operator for detection and precise location of distinct points, corners and centers of circular features. In: Proc. of the ISPRS Intercommission Workshop on Fast Processing of Photogrammetric Data, pp 281–305

Wehr A, Lohr U (1999) Airborne laser scanning an introduction and overview. ISPRS Journal of Photogrammetry and Remote Sensing 54(2-3):68–82

Wei M, Schwarz KP (1990) Testing a decentralized filter for gps/ins integration. Proceesdings of the IEEE PLANS - Position, Location and Navigation Symposium p March 1990

Weiss J, Kee D (1995) A direct performance comparison between loosely coupled and tightly coupled gps/ins integration techniques. In: in Proceedings of the 51st Annual Meeting of the Institute of Navigation June 5 - 7, 1995, Colorado Springs, CO, pp 537 – 544

Wolf P (1974) Elements of photogrammetry. McGraw-Hill

Index