Data Acquisition in Geographic Information Systems 9

Jan Skaloud, Michael Cramer, and Norbert Haala

Contents

9.1 **Optical Sensors** . . . 185
9.1.1 Frame Cameras . . . 186
9.1.2 Line Sensors . . . 190
9.1.3 Lidars . . . 192

9.2 **Navigation Sensors** . . . 196
9.2.1 Mapping Prerequisites . . . 196
9.2.2 Satellite Navigation . . . 197
9.2.3 Inertial Navigation . . . 200
9.2.4 Integrated Navigation . . . 200
9.2.5 Geometrical Relations . . . 202

9.3 **Photogrammetry** . . . 204
9.3.1 From 2-D to 3-D . . . 204
9.3.2 Camera Pose in a Homogeneous Form . . . 204
9.3.3 Pinhole Camera . . . 205
9.3.4 Image Coordinates . . . 206
9.3.5 Imaging Formation Model . . . 207
9.3.6 Scene from Two Views . . . 208
9.3.7 Scene from Multiple Views . . . 210
9.3.8 Feature Matching . . . 211

9.4 **Sensor Fusion** . . . 215
9.4.1 Principle . . . 215
9.4.2 Parameters . . . 217
9.4.3 Optical Distortion Models . . . 218
9.4.4 Observation Models . . . 218
9.4.5 Estimation . . . 220
9.4.6 Adopted Approaches . . . 221

9.5 **Mapping Products** . . . 224
9.5.1 Surfaces . . . 224
9.5.2 Orthophotos . . . 227

References . . . 228

J. Skaloud (✉)
Swiss Federal Institute of Technology Lausanne (EPFL)
Lausanne, Switzerland
e-mail: jan.skaloud@epfl.ch

M. Cramer
Institute for Photogrammetry (ifp), Stuttgart University
Stuttgart, Germany
e-mail: michael.cramer@ifp.uni-stuttgart.de

N. Haala
Institute for Photogrammetry (ifp), University of Stuttgart
Stuttgart, Germany
e-mail: norbert.haala@ifp.uni-stuttgart.de

Abstract

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 together with Lidar that serves the acquisition of detailed 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. The geometrical principals of 3-D restitution of the scene are described first in Photogrammetry for the case of frame imagery only, later in Sensor Fusion for active sensors and integrated approaches. An overview of Mapping Products concludes this chapter.

Keywords

optical sensors · navigation sensors · photogrammetry · sensor fusion · mapping · adjustment · Lidar · INS/GNSS

9.1 Optical Sensors

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 photosensitive materi-

© Springer Nature Switzerland AG 2022

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_9

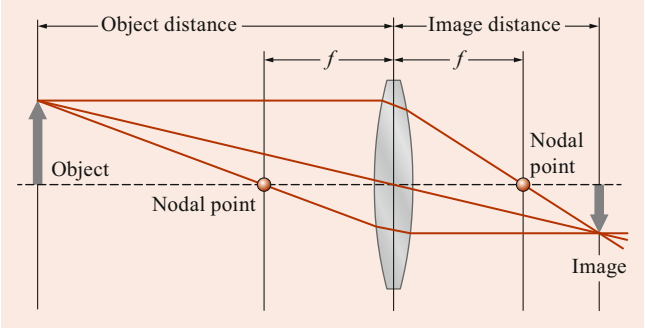

Fig. 9.1 Basic imaging principe when using a lens or lens system

als or elements. 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, principally, has a faster readout.

Photography in its simplest case is based on the pinhole camera model [1]. The geometric theory of optical systems assumes straight light rays that have been reflected from an object illuminated by any light source. These rays enter 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 Sect. 9.3.3, 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 defocus. It is assumed that pinhole, object point, and image point define 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 forms 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 a 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 pass the lens (system) without deviation, forming the same angles to the optical axis in both nodal points [2], i.e., the emerging ray is parallel to the original incident ray (Fig. 9.1).

The nodal points define the object–space incident and image-space emergent perspective centers. Light rays entering the lens parallel to the optical axis come into 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 [1]. Any parallel rays entering the system come into 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 Sect. 9.3. Reconstruction of 3-D coordinates of objects derived exclusively from 2-D image observations is detailed in Sect. 9.3, while the benefit of integrating navigation sensors into this process is described in Sect. 9.4.

9.1.1 Frame Cameras

Until recently, the acquisition of texture was done exclusively using films. These have now been almost completely replaced by electronic sensors, at least for consumer-grade photography. Consequently, and similarly to consumer grade photography, analog film-based systems are rapidly being phased out in operational photogrammetric environments, and digital sensors have, to a large extent, replaced their analog predecessors. According to their geometry 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 concepts of both of these are described in this and the following Sect. 9.1.2.

Single Head

Airborne or satellite platforms employ 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 analog mapping cameras have been designed with large formats of about $23 \times 23\,\text{cm}^2$. For these, focal lengths of 30, 15, or 8 cm are utilized, depending on the field of view (FOV) needed, which is 60, 95, and 125 degrees, respectively.

Multiple-Head Cameras

The most intuitive way to design a digital mapping camera would be to replace the former analog film by a 2-D electronic sensor element or sensor matrix. Indeed, such an approach has been pursued in consumer-grade photography.

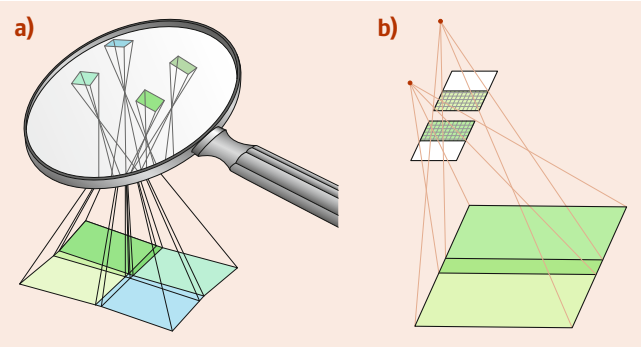

Fig. 9.2 Camera concepts based on CCD frame arrangements, synchronous recording: (**a**) "butterfly arrangement" with 4 tilted heads, each head with one CCD located in center position, (**b**) "shifted CCD arrangement", here with 2 nadir-looking heads, each with one CCD located in shifted position. Courtesy of ifp-Stuttgart

Unfortunately, the size of a CCD frame is physically limited by the supporting electronics. Therefore, was some years until a special multiple-head concept was developed based on a cluster of CCD sensors with a format comparable to the former 35 mm format (24 mm × 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, individual CCDs of a smaller format connected to separate camera heads generate multiple smaller-format images with certain overlaps. This allows for the generation of one synthetic large-format image afterwards, which is obtained by resampling 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, multihead 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, the so-called butterfly pattern, which is necessary to form a virtual image of a large format (Fig. 9.2). These overlaps are necessary for the later transfer of corresponding points (also called tie points) or tie features, which 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 decentered, shifted to the direction of the 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 (Fig. 9.2).

Syntopic Frame

In the concepts of multihead frame cameras presented so far, 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 cameras takes images at different times, however, over the same place. Such an 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. 9.3). If the different camera heads take their images one after the other, and this time shift corresponds exactly 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. Depending on the individual arrangement of CCD frames within the different camera heads, overlapping images are generated in the object space, which can again be merged together afterwards. The concept of image formation from syntopic imaging is illustrated in Fig. 9.3. As can be seen, up to four CCDs are placed within one single focal plane in a special pattern. The cone containing the four CCD sensors

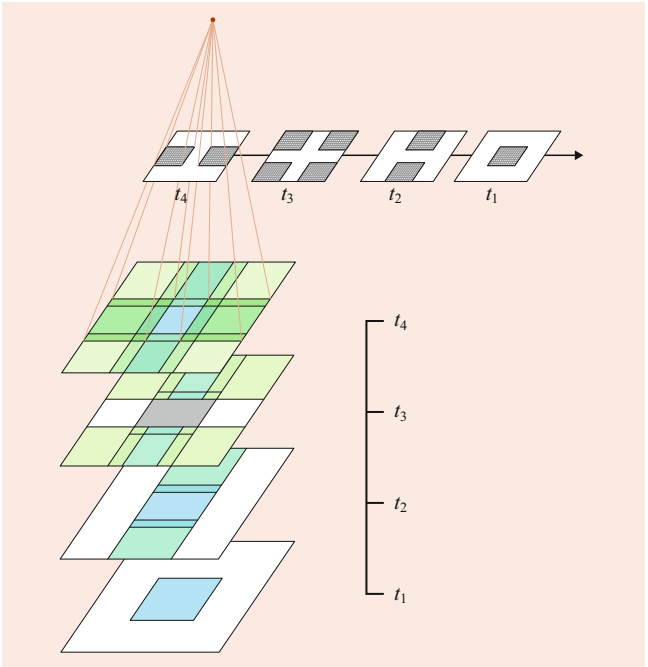

Fig. 9.3 The syntopic imaging concepts. Courtesy of ifp-Stuttgart

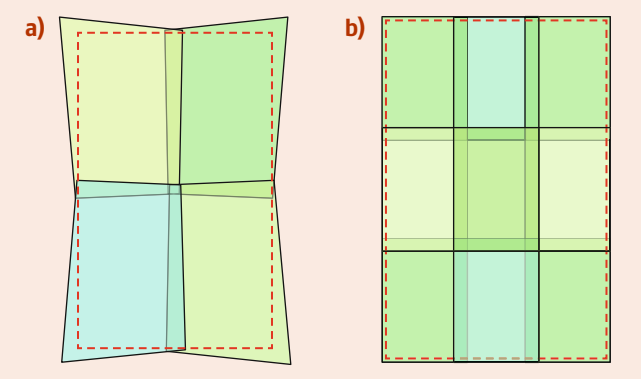

Fig. 9.4 Multihead concept to generate large-format frames: synchronous imaging using 4 tilted camera heads (**a**) and syntopic imaging using 4 nadir-pointing camera heads (**b**). Courtesy of ifp-Stuttgart

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.

Comparison of Concepts

Figure 9.4 compares the two ground patterns of the two main multihead 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. Fig. 9.4a shows the particular footprint of such a 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 influences the imaging of the 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 [3].

Syntopic image recording delivers a different pattern (Fig. 9.4b). Again, the color shades indicate the arrangement of CCD frames in the four participating camera cones. All cameras record 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. 9.4 indicates those parts of the images that 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 the 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.

Virtual Frame

The multihead concepts allow for the generation of virtual images of a larger format. For this purpose, several individual images are resampled 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 intercone orientation or *image stitching*. The knowledge of interior orientation (see Sect. 9.4 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 six 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 camera modules is derived from conjugate points measured from the overlaps between the different images [3, 4]. Even in multihead 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.

Color Generation

The generation of multispectral 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 to employ separate camera heads for each of the requested multispectral channels. Appropriate filters let each CCD array only capture the corresponding color information. Red, green, blue, and additional near-infrared spectral bands are the 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 colors. 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.

Figure 9.5 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 sen-

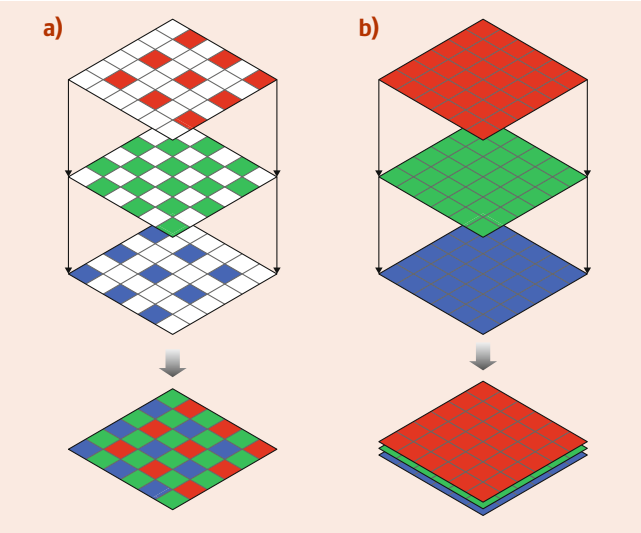

Fig. 9.5 The two concepts to obtain colored RGB digital images from frame sensors: (**a**) single CCD using the Bayer pattern. (**b**) individual CCDs, one for each color channel

sitive 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 head per color band or, alternatively, a beam splitter in the camera optic to separate the different color bands within one single camera head.

Color Resolution

Multiple-head cameras usually separate registration of panchromatic (gray values) and color channels. As color is often generated using the Bayer pattern, this results in a somewhat lower geometric resolution with respect to the panchromatic (PAN) image. Also, the design of multispectral channels is likely to set even lower spatial resolution compared to the large-format virtual pan image. High-resolution color imagery is then obtained from postprocessing, where the lower-resolution color channels are combined with the high-resolution PAN images. This process is called pansharpening and is frequently used in satellite imaging. The ratio between the spatial resolution of the pan and color channels is called the pansharpening ratio. Different approaches are used for pansharpening [5]. The methods can be classified into substitution approaches and arithmetic and filter-based techniques. The preservation of the original radiometric color information, depending on the algorithm is exemplarily discussed in [6]. The basic idea of the pansharpening concept is illustrated in Fig. 9.6. This example is taken from digital airborne image data. After the fusion of the lower-resolution RGB image (Fig. 9.6b) with the higher-resolution PAN image (Fig. 9.6a), pansharpening delivers a color image with higher geometric resolution of the PAN channel (Fig. 9.6c). In this example, one low-resolution RGB pixel corresponds to 4×4 high-resolution PAN pixels, which equals a $1:4$ pansharpening ratio.

Fig. 9.6 The concept of pansharpening to increase the spatial resolution of color imagery. (**a**) High-resolution PAN, (**b**) RGB low resolution, and (**c**) RGB after pansharpening. Courtesy of ifp-Stuttgart

9.1.2 Line Sensors

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 a few CCD lines are 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) recording almost continuously. This line scanner concept is also called 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, 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, then 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 pushbroom image data (Sect. 9.4). The pixel size obtained 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-track and across-track are independent, quadratic pixels 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

$$\begin{aligned} \mathrm{GSD}_{\mathrm{along}} &= v\Delta t \\ \mathrm{GSD}_{\mathrm{across}} &= \Delta y m_b = \Delta y \frac{h_{\mathrm{g}}}{f} , \end{aligned} \tag{9.1}$$

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, and Δy the pixel size across the flight direction.

Typically, more than one CCD line is used in a line scanner system (Fig. 9.7). 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 multispectral imaging. Additional lines are simply placed in the same focal plane that is already used for the panchromatic channels. Often, at least three panchromatic channels, as well as four multispectral channels, are used. All CCD lines provide the same number of pixels. Thus, panchromatic and multispectral images are obtained with the same geometric resolution, which is again a characteristic different from those of frame-based sensors. Even though almost all systems have more than three lines, such pushbroom systems are often referred to as three-line scanners. This is named after the three panchromatic lines.

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. 9.7 three multispectral channels are exemplarily placed in the nadir-viewing direction plus

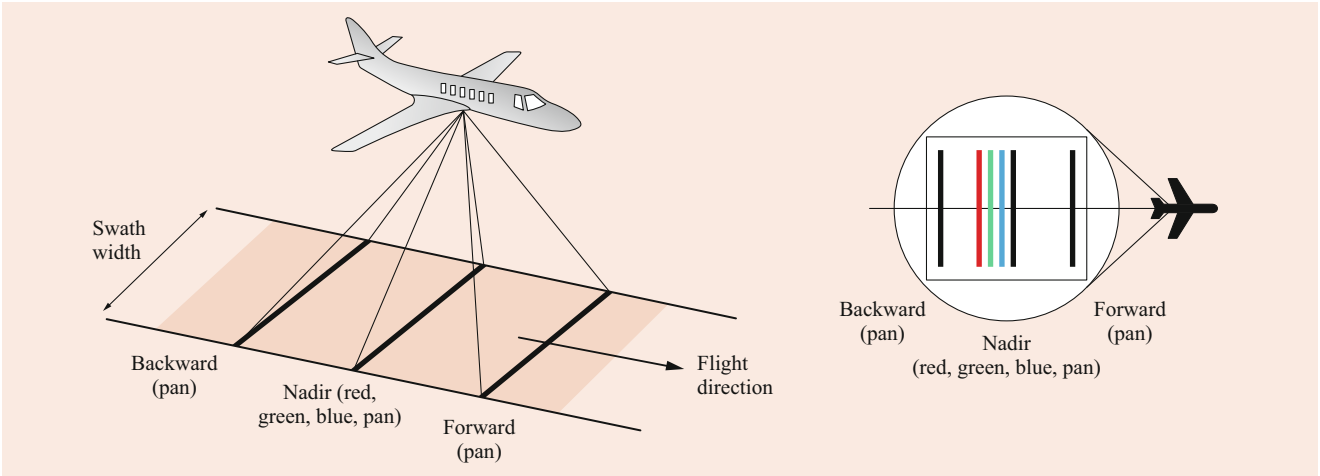

Fig. 9.7 The concept of an airborne line scanning system, after [7]

the three additional panchromatic lines in forward, backward, and nadir directions. Thus, three different stereo angles are possible, namely between the forward and nadir, backward and nadir, and forward and backward views. 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 multispectrum stereo viewing capabilities.

As mentioned, the stereo capability depends on the different viewing directions due to the parallax effect. Nevertheless, such parallaxes also appear in the different multispectrum lines, which are needed to later obtain the full color images by combination of individual spectral bands. Even though these color/multispectral lines are typically mounted as closely as possible, their displacement will cause different perspective distortions in each spectral band. The larger the distance between the different spectral CCD lines, the larger the influence of these displacements. 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 (Sect. 9.5.2). 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 an installation allows us to exactly coregister the four different color bands (Fig. 9.8). Thus, 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.

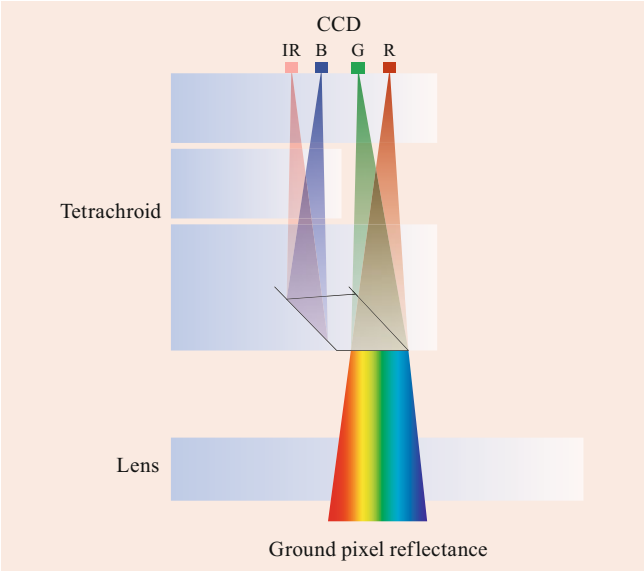

Fig. 9.8 Beam splitter used in a pushbroom sensor (© Leica Geosystems)

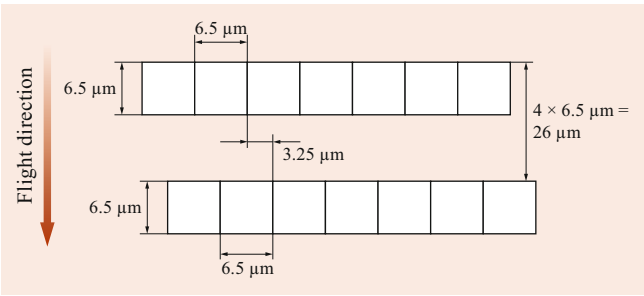

Fig. 9.9 Staggered line arrangement, situation in the focal plane

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 12 000 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. 9.9 shows the arrangement of a staggered line with $6.5 \times 6.5\,\mu m^2$ 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 2. 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.

Figure 9.10 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. 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 area flown over are imaged at the same time. In this example, line A at time t_2 covers the same area that was already 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

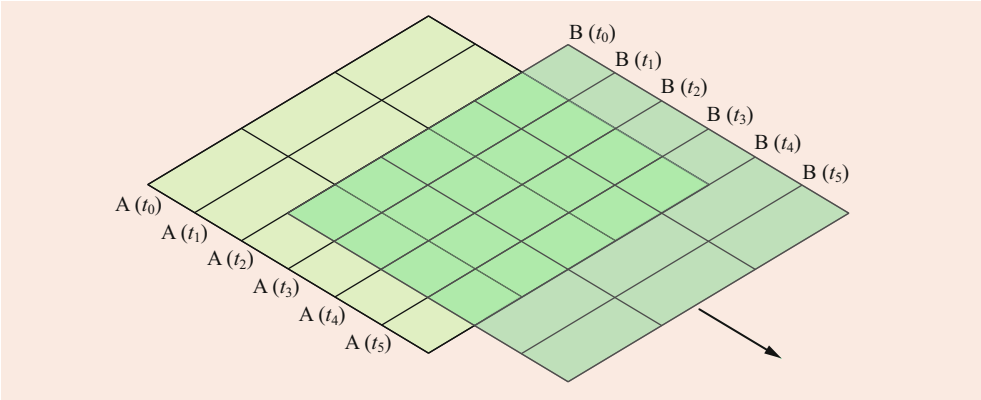

Fig. 9.10 Sampling pattern on the ground with the concept of line staggering. Courtesy of ifp-Stuttgart

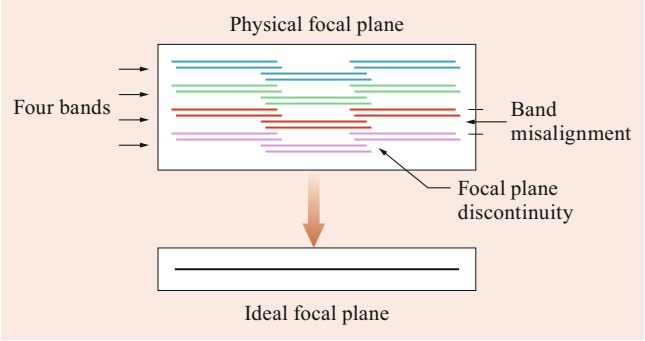

Fig. 9.11 Stepped arrangement of multiple CCD lines

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 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-borne sensors, where the trajectory is much smoother compared to airborne platforms [8].

Another approach to increase resolution and swath width is the employment of multiple CCD lines that are shifted against each other across the flight direction. Since these CCD lines cannot be stitched together directly, a stepped arrangement is necessary (Fig. 9.11). Although this approach has so far been 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.

9.1.3 Lidars

The acquisition of terrain structure is very efficiently achieved by optical sensors such as radar and lidar. Both methods are active, i.e., an energy is emitted from the sensors, and 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.

Laser Ranging

Introduced towards the end of the last millennium, lidar 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 with 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. 9.12). Hence, the distance ρ from the ranging unit towards the target is deduced by the following relation

$$\rho = \frac{1}{2} c\, t \ , \tag{9.2}$$

where c is the speed of light. As shown in Fig. 9.12, the laser-ranging unit comprises an emitting laser and an electro-

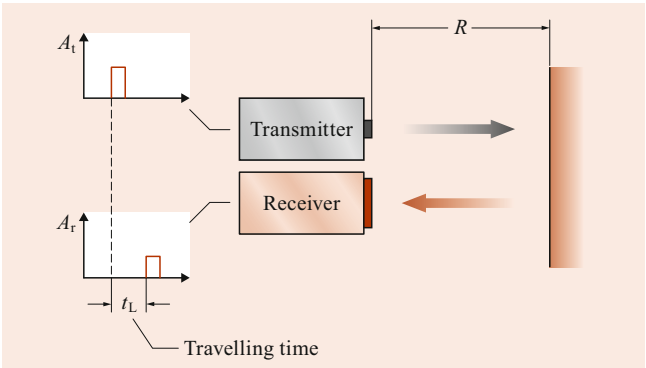

Fig. 9.12 Lidar ranging principle; A_t and A_r are the amplitudes or intensity of the transmitted and received pulses, respectively, after [9]. Courtesy of ISPRS

optical receiver. The transmitting and receiving apertures are oriented in the same direction to ensure that the system will detect the target to which the transmitter points. 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 0.1–3 mrad.

There are two technological principles of laser ranging that are implemented in mapping applications: continuous wave (CW) lasers and pulse lasers. In CW lasers, the radiation is emitted as a continuous beam instead of a sequence of discrete pulses. This limits the power of the CW laser to terrestrial laser scanning, although there are exceptions [10]. 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

$$t = \frac{\phi}{2\pi} T + nT \ , \tag{9.3}$$

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 [9].

Although CW lasers reach higher-ranging accuracies, today's airborne lidar systems almost solely use pulsed ranging. A pulse laser functions as shown in Fig. 9.12. The widespread of pulse lasers is due to two technological advances. First, the progress in accurate quartz-stabilized oscillators enables determining the time elapsed 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 [11]. In long-range (airborne) applications, the different implementations of pulse-based rangefinders can be distinguished:

- *Linear mode–discrete echo*. After emission of a 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 us to scan even through the canopy, because the spacing between the leaves and branches allows parts of the pulse to penetrate further into the ground, while some energy is reflected immediately. This principle is schematically depicted in Fig. 9.13. As shown in the third plot of Fig. 9.14, the partial reflections are detectable above a 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 analog echo waveform, i.e., the time-dependent variation of received signal power, for each emitted laser pulse (lowest plot in Fig. 9.14). 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 inter-

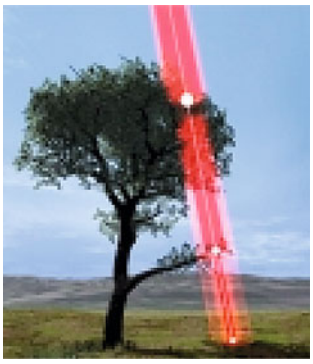

Fig. 9.13 Principle of multiple echoes from a laser signal. After [12], courtesy of the author

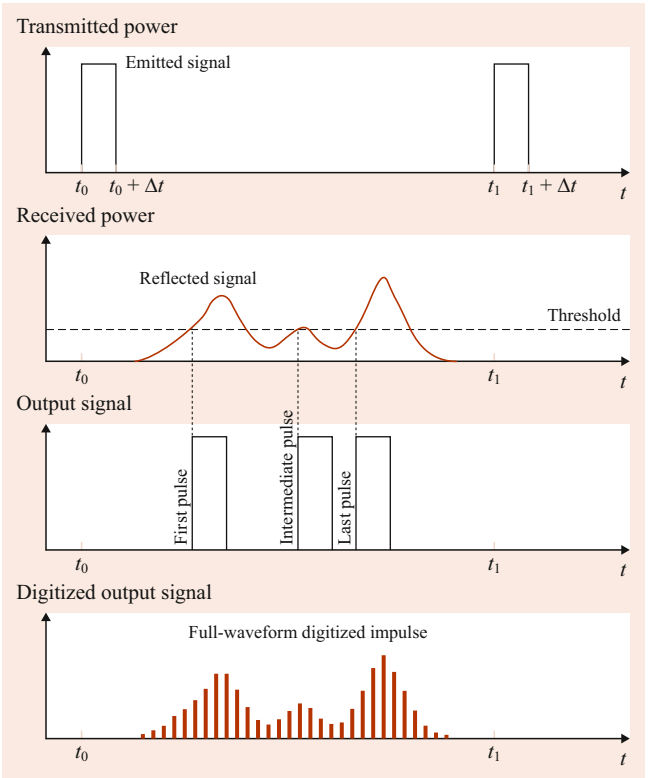

Fig. 9.14 Emitted and received impulse for discrete echo scanners and full-waveform scanners. After [12]

val 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" counter requires a hundred times less power to register a return than linear mode lidar detectors. The large sensitivity of Geiger detectors allows considerably longer ranging than that of linear mode scanners. Coupling long-ranging capability with the employment of a large number of small detectors and a high repetition rate (hundreds of MHz) allows maintaining a few pulses per m^2 from 5 to 10 km above ground, which increases the swath width and, thus, the productivity considerably. However, the first generation of these detectors allows us 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 events 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 than Geiger-mode scanners, albeit not yet as precise as linear-mode lasers. The first commercially available ALS of this type was introduced in 2018. In 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) wavelength, while single-photon lasers currently use 530 nm (green lasers). 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 manmade structures, whereas surfaces with water content (i.e., glaciers, snow) reflect weakly. On the other hand, 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 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 sunlight 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.

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 a vehicle. As is schematically shown in Fig. 9.15, the 2-D 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 a carrier's position was difficult to achieve, which was the reason that laser altimetry was used almost exclusively on spaceborne platforms. There, the motion was determined by satellite-tracker observations and by

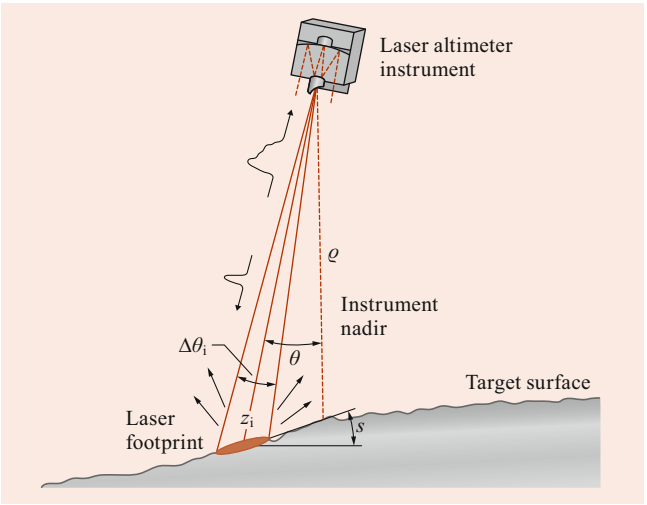

Fig. 9.15 Lidar profiling from a spaceborne platform using a laser altimeter

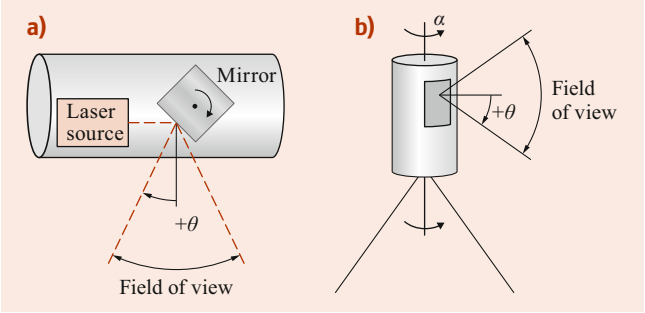

Fig. 9.16 Lidar scanners: (**a**) 2-D laser scanner with a rotating polygonal mirror, (**b**) 3-D scanner based on the same principle

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, 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), [13]) 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 [14], 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 [15].

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 2-D elevation profile Δh with respect to the leveled instrument is obtained as

$$\Delta h = \rho \sin(\theta) \tag{9.4}$$

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 2-D laser scanner which is described in the following section (Sect. 9.1.3, *Scanners*).

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 2-D or 3-D profiles, respectively. Frequently, the 2-D scanning mechanism is used (Fig. 9.16a), from which the 3-D 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. 9.16b)
- 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 $\phi_{\max}$ as

$$\mathrm{SW} = 2h \tan \frac{\phi_{\max}}{2} \,. \tag{9.5}$$

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. 9.17 and their comparison is provided in Table 9.1.

The potential of employing laser ranging for navigation and collision avoidance systems initiated the development of devices operating over shorter distances ($< 100\,\mathrm{m}$) without the scanning mechanism. There, a few tens of lasers are arranged in a line array with a regular angular separation and

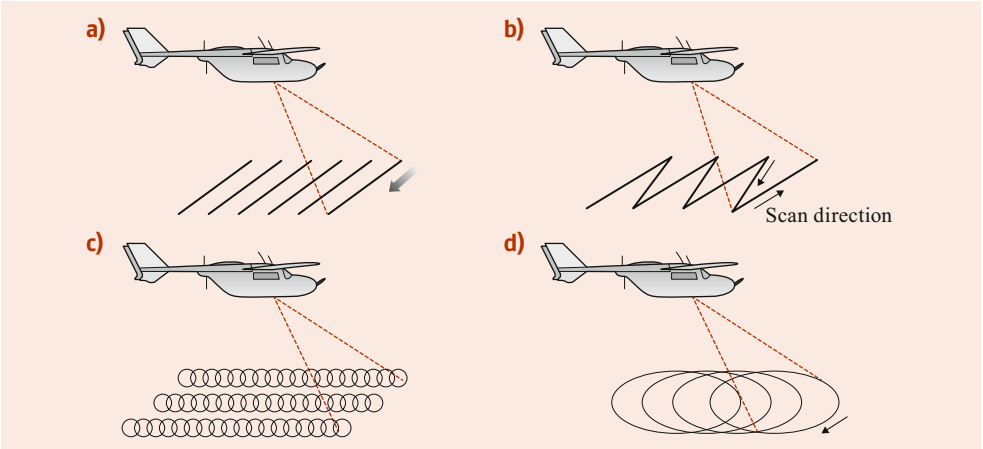

Fig. 9.17 Different scanning patterns (after [16]): (**a**) polygonal mirror, (**b**) oscillating mirror, (**c**) fiber scan, and (**d**) Palmer scan

Table 9.1 Comparison of different scanning patterns used in mobile laser scanning

Mechanism	Characteristics	
	Advantages	Disadvantages
Polygonal mirror	• Constant rotation avoids mirror distortions due to additional force • Provides regularly spaced sampling along and across 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 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 torsion • Z-shaped irregular sampling with lower density at nadir
Fiber 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 due 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

FOV of 30°. The 3-D profile is created by rotating the whole assembly, similarly to in Fig. 9.16b, nevertheless with rotation rates up to several tens of Hz, resulting in a high data collection rate. Although the ranging is generally less precise than for scanning lasers, after proper calibration [17, 18], these devices have applicability in mapping from ground vehicles and UAVs.

Scanner frames

The definition of a scanner frame is chosen arbitrarily and, therefore, differs among manufactures. The following definition applies to several systems and can be ported to other instruments by a simple permutation of axes. The location of a point within the scan line j can be conveniently expressed either by polar or Cartesian coordinates, with the former usually being used. Considering the situation depicted in Fig. 9.18, the relations between the *range measurement* ρ, the encoder *horizontal angle* θ, and the *vertical angle* α with respect to the scanner frame defined in Cartesian coordinates are

$$\rho_{ij} = \sqrt{x_{ij}^2 + y_{ij}^2 + z_{ij}^2}\,, \tag{9.6}$$

$$\theta_{ij} = \arctan\left(\frac{y_{ij}}{x_{ij}}\right), \tag{9.7}$$

$$\alpha_{ij} = \arctan\left(\frac{z_{ij}}{\sqrt{x_{ij}^2 + y_{ij}^2}}\right). \tag{9.8}$$

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

$$\boldsymbol{x}^s = \rho \begin{pmatrix} 0 \\ \sin\theta \\ \cos\theta \end{pmatrix}. \tag{9.9}$$

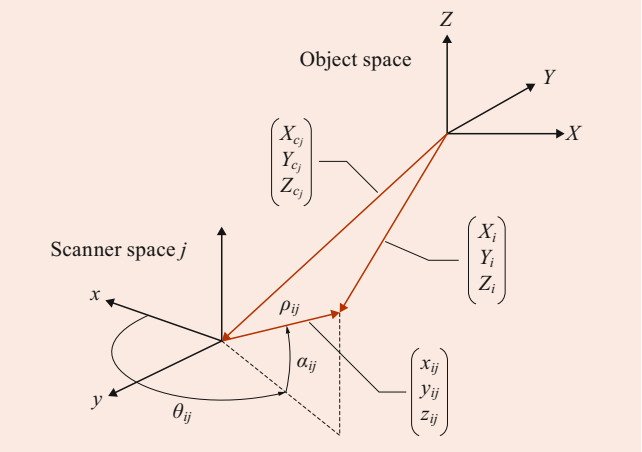

Fig. 9.18 Scanner frame and observation geometry

9.2 Navigation Sensors

9.2.1 Mapping Prerequisites

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

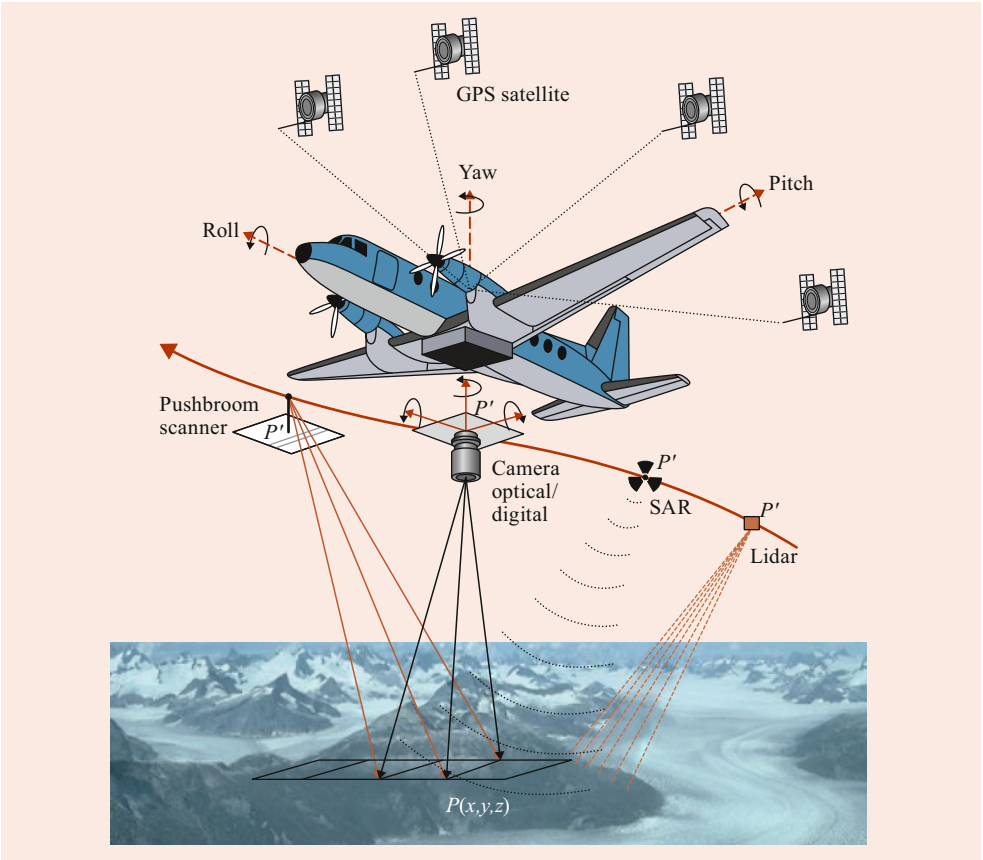

Fig. 9.19 Use of navigation technology for sensor orientation

geocoding, or *(sensor) orientation* and concerns the following components (Fig. 9.19):

- 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, Sect. 9.4). 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 differ for every object point. The following text provides an overview of the navigation technology that facilitates the problem of sensor orientation tremendously.

Direct measurement of EO parameters typically relies on integrating receivers of the Global Navigation Satellite System (GNSS), such as GPS (Global Positioning System), with an inertial navigation system (INS) with the possible aid 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 introduced separately, while their integration will be described in Sect. 9.2.4). The end of this section is devoted to the introduction of the reference frame and the establishment of relations for transferring the trajectory observation to sensors.

9.2.2 Satellite Navigation

Available Systems

Satellite navigation have global or regional character (Fig. 9.20). There are four global navigation satellite systems

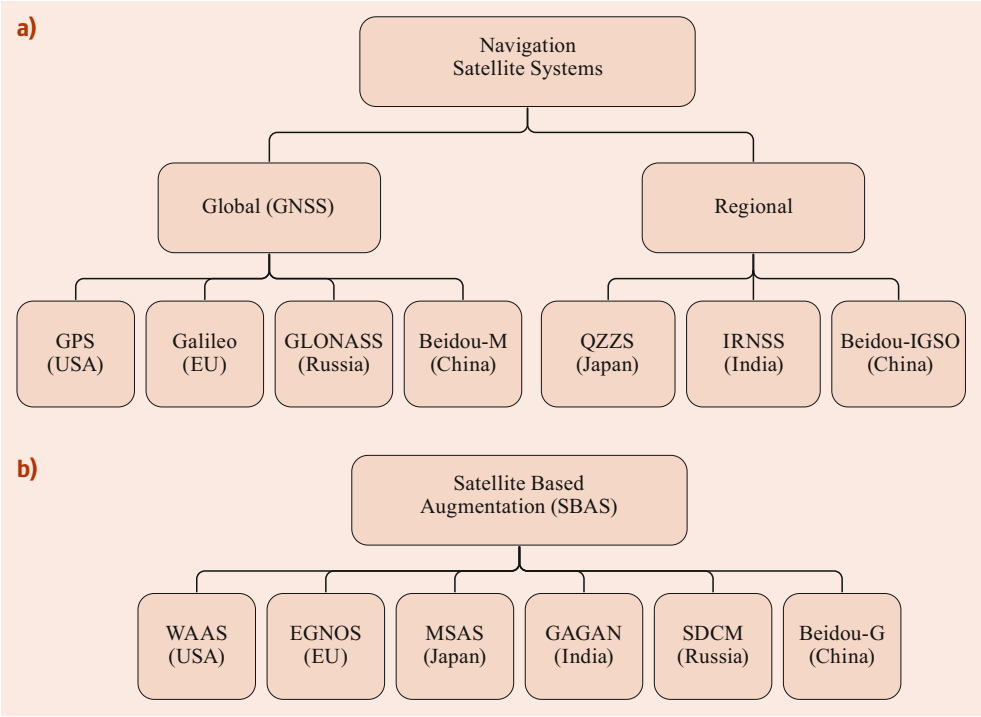

Fig. 9.20 (**a**) Overview of today's GNSS. *Left branch*: satellites on global orbits. *Right branch*: satellites moving only above regions. (**b**) Satellite-based augmentation system with regional implementation of stationary satellites

Table 9.2 GNSS nominal constellation characteristics, after [19]

Characteristics	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	19,130	21,400
Orbital period (hh:mm:ss)	11:58:02	14:04:41	11:15:44	12:52:04

(GNSS), in reverse 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 6 (GPS) or 3 (others) orbital planes at medium Earth orbit (MEO). The slight differences in orbital altitude among constellations result in different orbital periods, as denoted in Table 9.2. 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.

Regional satellite navigation employs Inclined Geosynchronous Orbit (IGSO) with nine (QZZS), seven (IRNSS), and three (Beidou-IGSO) satellites, respectively. In addition to emitting proprietary navigation services, these regional satellites also broadcast 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. Because it used exclusively geostationary Earth orbits (GEO), SBAS is practically reaching a global coverage (bottom part of Fig. 9.20). 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 from 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.

Signal 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 (centered at 1575.42 MHz) and L2 (centered at 1227.60 MHz), while modernized GPS also includes L5 (centered at 1176.45 MHz). The signal that remains open to all users on all generations of GPS satellites is the coarse-acquisition (C/A) code transmitted on L1. The later generation of GPS satellites emit an additional open signal (L2C) on L2 frequency, while modernized GPS added an open L5 signal on a third frequency and an open L1C sig-

Table 9.3 Open signals of modern GNSS based on CDMA

GNSS	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

nal on the first frequency (Table 9.3). 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 such as multipath, while improving interoperability with Galileo and other systems.

To distinguish signals coming from different satellites GPS, Galileo and Beidou adopted code division multiple access (CDMA), while GLONASS used frequency separation (FDMA), which made the fabrication of receivers more complex. To improve the interoperability with other systems, modernized GLONASS added CDMA on three frequencies while keeping FDMA for continuity. As shown in Table 9.3, the open service with CDMA on GLONASS is, however, only available on two frequencies. The situation is somewhat similar for Galileo and Beidou, which both adopted complex message structures on E5 and B2 that resulted in a 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. Due to the interoperability among GNSS the number of available satellites increased substantially over 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 further improve the accuracy and reliability of satellite-based positioning.

Positioning Methods

An overview of current GNSS positioning techniques is provided in Fig. 9.21. The selection of a particular method depends on the factors of accuracy and rapidity in data acquisition and mobile mapping. These methods are the following:

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

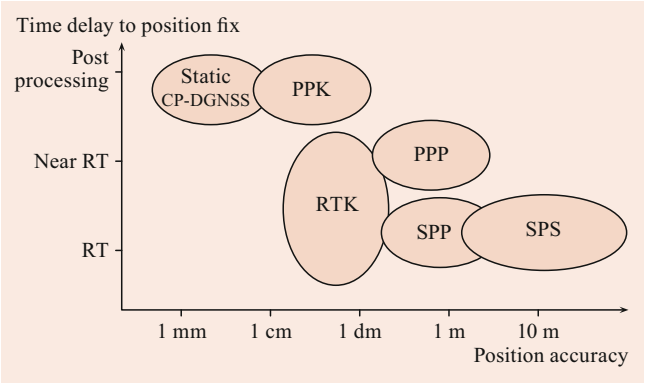

Fig. 9.21 Overview of GNSS positioning methods as a function of accuracy and rapidity

- *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. This technique can achieve subdecimetric position accuracy [20] and is available worldwide without the need for an augmentation system.
- *Differential GNSS (DGNSS)*, *carrier-phase DGNSS* and *postprocessed kinematic (PPK)* are relative positioning techniques based on simultaneous observations by the rover and base (one 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 also employ 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, subcentimeter to millimeter accuracy can be achieved when respecting some considerations about baseline length and observation time. The upper limit in PPK is centimeter to decimeter accuracy for a 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 decimeters, while ambiguity resolution using observations on a single frequency is limited to a 1–2 km long baseline.
- *Real-time kinematics (RTK)* applies the above-mentioned PPK principles in real time. 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 types 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. Subdecimeter to centimeter-level positioning accuracy can be achieved by these means in ideal conditions.

9.2.3 Inertial Navigation

Inertial navigation derives position, velocity, and attitude from the initial knowledge of these quantities and from the integration of the accelerations observed (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 in 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 [21, 22]; the following list is limited to the types common in direct sensor orientation.

Gyroscope Technology

Gyroscopes usually represent the most expensive part of an IMU. Their accuracy significantly affects 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 the 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 [23]. 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, but 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, which produces small and inexpensive sensors. They are used in mass market, automotive, robotic, and entry-level navigation applications. They are indispensable on drones for control and guidance. As their high end in some aspects approaches the quality of low-end FOGs, they are useful for robotics and UAV-based acquisition.

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.

Strapdown INS

In earlier INS realizations, 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, inertial sensors are rigidly mounted (strapped down) to their casing, thereby 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 (Sect. 9.4.6).

9.2.4 Integrated Navigation

Principle

Integrated navigation is a technique that combines data from several navigation systems or sensors with the aim of improving the accuracy and robustness of the estimated trajectory. In this respect, satellites 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 their 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 of the 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, inertial navigation provides continuous data output for all trajectory parameters (i.e., position, velocity, and attitude). Therefore, the integrated navigation principally stabilizes and refines the 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 on external infrastructure or visibility. Navigation of this type may be provided by odometers, pressure sensors, Doppler radar, or magnetic sensors, for example.
- *Dynamic constraints*: the application of implicit knowledge of some dynamical state or its form. For example, constraints such as zero velocity and nonholonomic condition (i.e., the alignment of vehicle speed with its direction) are used as a supplementary aiding method on terrestrial vehicles or complete vehicle dynamic model is used for UAVs [24, 25].

Integration Schemes

Optimal integration of different measurement data with inertial observations is commonly achieved by using a Kalman filter/smoother (Fig. 9.22). Data filtering/smoothing can, however, be organized in different manners with respect to GNSS observations. The following two integration schemes are the most important:

- *Loosely-coupled integration*. This is the most common integration approach, especially in airborne or shipborne installations. Raw IMU measurements are integrated to yield position and attitude at the IMU output rate (normally 100–500 Hz). The position and velocity data gathered by GNSS are processed independently, yielding a sequence of positions and velocities at a certain frequency (normally 0.1–1 Hz). These data are subsequently fed as updates in an extended Kalman filter (EKF). The differences observed between the predicted (INS) and GNSS-determined velocities and positions are used to estimate the elements of the filter state vector, containing, on one hand, the error states related to the trajectory (i.e., position, velocity, and orientation) and, on the other hand, 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, 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 reception (e.g., urban canyons), and is commonly used in terrestrial mobile mapping.

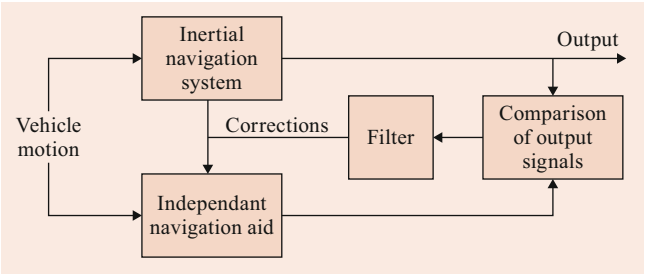

Fig. 9.22 Typical integrated navigation scheme for direct georeferencing

A performance comparison between the two integration schemes presented can be found in [26, 27] and, with a focus on RTK, in [28].

Resulting Accuracy

For strapdown INS, sensor integration solves firstly the problem of calibrating the systematic errors (i.e., residual gyro and accelerometer biases, scale factors, etc.). Secondly, the use of GNSS data mitigates attitude initialization 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. (This concerns all gyroscopes of lower accuracy, like those employed in UAVs that cannot complete north-seeking without an external assistance.) At the same time, the inertial system smoothens noisy velocity outputs from GNSS and provides high-rate measurement of position and velocity over a larger spectrum of motion.

There is no such thing as a perfect instrument, and as strong as it is, the integration cannot eliminate all errors completely. In other words, the data integration handled by a Kalman filter/smoother only cancels the nonoverlapping part of the sensor's error budget. The performance of error cancelation depends on the motion of interest, the instrument type, and the encountered dynamics. While the long-term positioning accuracy limit depends on the GNSS positioning

Table 9.4 Orientation accuracy of as a function of time and INS quality

Time	Navigation grade		Tactical grade	
	Roll/pitch (°)	Yaw (°)	Roll/pitch (°)	Yaw (°)
1 s	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 9.5 Orientation accuracy for small MEMS IMU of high quality

Time	Low-end tactical grade	
	Roll/pitch (°)	Yaw (°)
1 s	0.005–0.1	0.005–0.2
1–3 min	0.03–0.2	0.1–0.2
Longer time	Trajectory dependent	

solution (Fig. 9.21), the time over which such accuracy can be maintained in the absence of satellite signals mainly depends on the quality of the INS and its preceding calibration. Based on the position error accumulated after 1 h of autonomous operation, the INS are normally grouped into four main categories [29]: *strategic 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 Table 9.4. The automotive in Table 9.5 corresponds to small MEMS IMUs of high quality, like those used by terrestrial and indoor robots or UAVs.

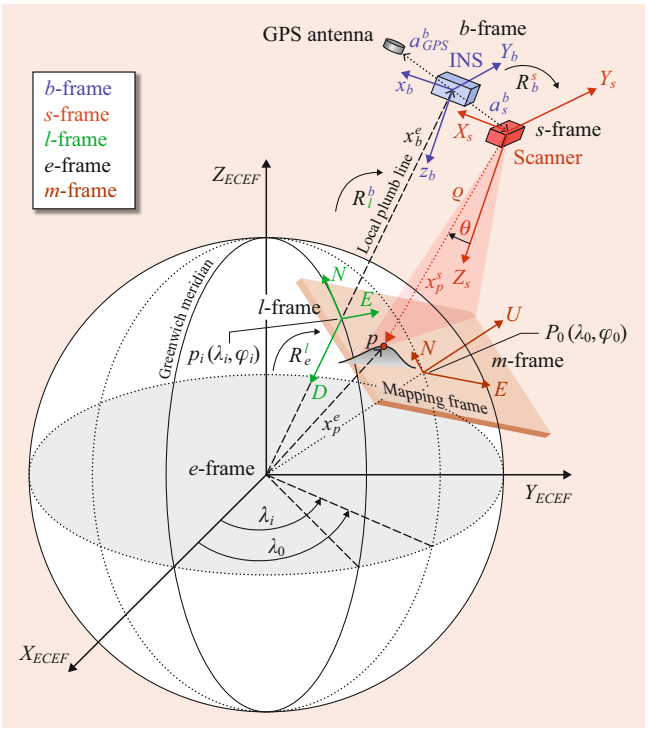

Fig. 9.23 Geometry of direct sensor orientation

9.2.5 Geometrical Relations

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 a global reference frame, as well as several intermediate frames. Table 9.6 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. 9.23 for the case of an airborne laser scanner. The situation is similar for other optical devices for which the sensor frames(s) are defined later (Sects. 9.1 and 9.4).

Earth-Centered, Earth-Fixed Frame – *e*

The satellite orbits of common GNSS systems are referred to this frame, and so is the outcome of the trajectory computation. A geocentric ellipsoid is normally attached to an ECEF frame, and its properties together with other geophys-

Table 9.6 Overview of reference frames

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 2-D 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's rotational 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

ical 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 parameterization is often used in the output of GNSS/INS trajectory. The relation between the Cartesian and ellipsoidal coordinates reads

$$\boldsymbol{x}^e = \begin{pmatrix} x_1^e \\ x_2^e \\ x_3^e \end{pmatrix} = \begin{pmatrix} (N+h)\cos\varphi\ \cos\lambda \\ (N+h)\cos\varphi\ \sin\lambda \\ \left(\frac{b^2}{a^2}N + h\right)\sin\varphi \end{pmatrix}, \tag{9.10}$$

where N is the radius of curvature in the prime vertical, and a and b are the semimajor and semiminor axes, respectively.

Local-level frame – *l*

This frame is mainly used as the reference for the orientation angle output from 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 downward to complete the system. Such a 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-frame to the e-frame can be described by the matrix $\mathbf{R}^e_{l_{\mathrm{NED}}}$

$$\mathbf{R}^e_{l_{\mathrm{NED}}} = \begin{pmatrix} -\sin\varphi\ \cos\lambda & -\sin\lambda & -\cos\varphi\ \cos\lambda \\ -\sin\varphi\ \sin\lambda & \cos\lambda & -\cos\varphi\ \sin\lambda \\ \cos\varphi & 0 & -\sin\varphi \end{pmatrix}. \tag{9.11}$$

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 points forward along the fuselage, the y^b-axis points to the right, and the z^b-axis points downward. 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

$$\mathbf{R}^b_{l_{\mathrm{NED}}} = \mathbf{R}_x(r)\,\mathbf{R}_y(p)\,\mathbf{R}_z(y)\,. \tag{9.12}$$

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

$$\begin{aligned} \mathbf{R}_x(r) &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos r & \sin r \\ 0 & -\sin r & \cos r \end{bmatrix} \\ \mathbf{R}_y(p) &= \begin{bmatrix} \cos p & 0 & -\sin p \\ 0 & 1 & 0 \\ \sin p & 0 & \cos p \end{bmatrix} \\ \mathbf{R}_z(y) &= \begin{bmatrix} \cos y & \sin y & 0 \\ -\sin y & \cos y & 0 \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned} \tag{9.13}$$

Transformation of the Exterior Orientation

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

$$\boldsymbol{x}^b_p = \boldsymbol{x}^b_{bs} + \mathbf{R}^b_s(\omega_b, \varphi_b, \kappa_b)\boldsymbol{x}^s_p\,, \tag{9.14}$$

where $\boldsymbol{x}^b_{bs} = \mathbf{R}^b_s \boldsymbol{x}^s_{bs}$ denotes the origin of the s-frame in the b-frame, which is also known as the lever-arm vector. The rotation matrix $\mathbf{R}^b_s$ in (9.14) is called the bore sight 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 bore-sight angles and the lever-arm vector 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. 9.23 by combining (9.10) with (9.12)

$$\boldsymbol{x}^e_p(t) = \boldsymbol{x}^e_b(t) + \mathbf{R}^e_l(t)\mathbf{R}^l_b(t)\mathbf{R}^b_s(\omega_b, \phi_b, \kappa_b)\left(\boldsymbol{x}^s_{bs} + \boldsymbol{x}^s_p(t)\right), \tag{9.15}$$

where $\boldsymbol{x}^e_b(t)$ 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 the ECEF-frame can be generated via (9.15). 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 pointwise or pixelwise via relations published by local surveying authorities. Alternatively, the registration of optical images and that of lasers can be performed directly in mapping frame as discussed in detail in [30, 31].

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 exist between optical and navigation sensors.

The lever arm $\boldsymbol{x}_{bs}^{b}$ 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 $\boldsymbol{x}_{ba}^{b}$, which is needed during GNSS/INS integration. Calibration of lever arms is best performed by tacheometry. Such a procedure is discussed in detail in [12, 32] for an aircraft and a small UAV system, respectively. An alternative solution is to estimate $\boldsymbol{x}_{ba}^{b}$ directly in the GNSS/INS Kalman filter/smoother as an additional parameter. Similarly, $\boldsymbol{x}_{bs}^{b}$ can be estimated in the block adjustment (Sect. 9.4), but its value is often strongly correlated to other parameters, and this approach should, therefore, be avoided when later used for direct sensor orientation.

The recovery of the bore-sight matrix $\mathbf{R}_{s}^{b}$ is more involved and requires the use of principles described in Sect. 9.4. For frame cameras, this process can be achieved in either one [33, 34] or two steps [35]. A similar procedure is maintained for line scanners [36, 37]. The bore-sight determination in kinematic laser scanning followed a rapid evolution [38–40] that converged to the approach based on surfaces of known form [41, 42]. The calibration principles are further addressed in Sect. 9.4.

9.3 Photogrammetry

9.3.1 From 2-D to 3-D

The main task of photogrammetry or, equivalently, *computer vision* is to reconstruct 3-D scenes from 2-D images. The most important requirement for the reconstruction to work is the that the scene be 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, which fact allows us 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), like those discussed in Sect. 9.2, can be used in support of the reconstruction process, which is a subject of Sect. 9.4.

Given a set of images, the principal challenges (and steps) of reconstructing 3-D models are threefold (Fig. 9.24):

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

This section essentially concentrates on the geometry part of the reconstruction problem without the help of navigation sensors. It is known under different names, such 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.

This chapter proceeds as follows. Although a general introduction to the relation between sensor-mapping frames was given in Sect. 9.2.5, we repeat it here but in a homogeneous representation of coordinates (Sect. 9.3.2), the form of which will be useful later on. Then, after introducing the basic geometry of the imaging system (Sects. 9.3.3 and 9.3.4), we describe a model of image formation for an ideal perspective camera (Sect. 9.3.5. With the necessary components, we introduce the reconstruction process for a stereo-pair of images (Sect. 9.3.6), which we later extend to multiple views (Sect. 9.3.7). We conclude the chapter with different processing strategies for filtering and optimization in scene reconstruction.

9.3.2 Camera Pose in a Homogeneous Form

Consider two Cartesian frames, where one is a mapping frame spanning the object space, and the other is related to a camera, viewing a scene at certain time t, to which belongs a point p. From the situation depicted in Fig. 9.25, it is clear that the coordinates of a point p with respect to the mapping frame m are simply the sum of the translation $\boldsymbol{x}_{c}^{m}$ of the origin of the frame c relative to that of the frame m and the vector $\boldsymbol{x}^{c}$ expressed in relation to the mapping frame m, which is $\mathbf{R}_{c}^{m}\boldsymbol{x}^{c}$, where $\mathbf{R}_{c}^{m}$ is the relative rotation between the frames

$$\boldsymbol{x}^{m} = \boldsymbol{x}_{c}^{m} + \mathbf{R}_{c}^{m}\boldsymbol{x}^{c} \,. \tag{9.16}$$

Every time the camera moves, its motion is captured by $\mathbf{T}_{c}^{m} = (\mathbf{R}_{c}^{m}, \boldsymbol{x}_{c}^{m})$ or more shortly by $\mathbf{T} = (\mathbf{R}, \boldsymbol{x})$ when the frames involved are clear of the context. It will become an

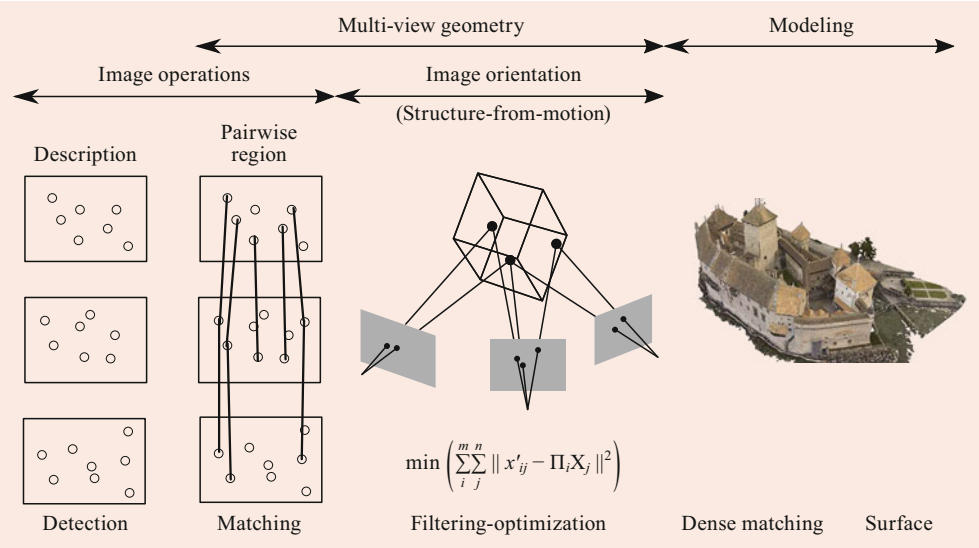

Fig. 9.24 Photogrammetry/computer-vision process for 3-D scene reconstructions based on 2-D imagery

advantage when we convert the transformation expressed by (9.16) to an expression of the form $\boldsymbol{u} = \mathbf{A}\boldsymbol{v}$. This is possible by adding "1" to the vector $\boldsymbol{x}$ as its fourth coordinate and by defining operations on so-called *homogeneous coordinates*. Such an extension preserves the original Euclidean space,

$$\overline{\boldsymbol{x}} \doteq \begin{pmatrix} \boldsymbol{x} \\ 1 \end{pmatrix} = \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} . \tag{9.17}$$

The vectors are defined analogically as differences of coordinates $\overline{v} = \overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2$. Differences makes the fourth component null and give rise to the original subspace. Rewriting (9.16) in the new notation leads to

$$\overline{\boldsymbol{x}}^m = \begin{pmatrix} \boldsymbol{x}^m \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{R}_c^m & \boldsymbol{x}_c^m \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^c \\ 1 \end{pmatrix} \doteq \overline{\mathbf{T}}_c^m \, \overline{\boldsymbol{x}}_c \, , \tag{9.18}$$

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

$$\overline{\mathbf{T}}_c^a = \overline{\mathbf{T}}_b^a \, \overline{\mathbf{T}}_c^b = \begin{pmatrix} \mathbf{R}_b^a & \boldsymbol{x}_b^a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{R}_c^b & \boldsymbol{x}_c^b \\ 0 & 1 \end{pmatrix} . \tag{9.19}$$

As can be easily verified, the inverse transformation is

$$\overline{\mathbf{T}}^{-1} = \begin{pmatrix} \mathbf{R} & \boldsymbol{x} \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{R}^\mathsf{T} & -\mathbf{R}^\mathsf{T}\boldsymbol{x} \\ 0 & 1 \end{pmatrix} . \tag{9.20}$$

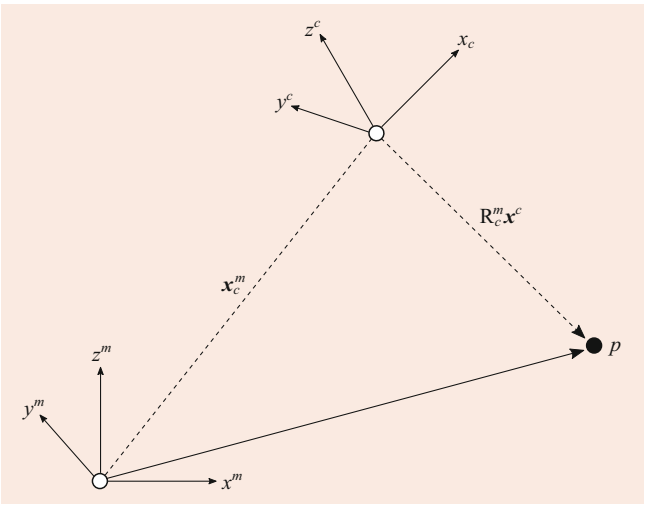

Fig. 9.25 Motion of a camera frame with respect to a Cartesian mapping frame

9.3.3 Pinhole Camera

Consider the basic imaging system as described by Fig. 9.1 in Sect. 9.1; the aperture of the lens decreases to zero, and the only feature that contributes to the 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. 9.26, where, for simplicity, the camera frame is oriented in the same direction as the object frame, and the point p is such that its image coordinate $y' = 0$.

Let us 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

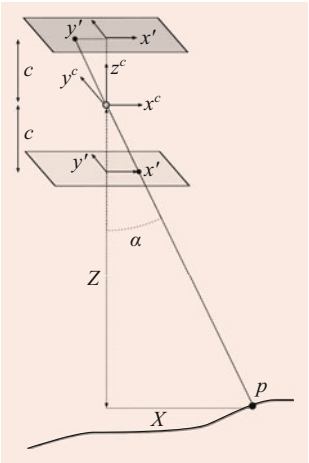

Fig. 9.26 Pinhole imaging model and its frontal counterpart: the 3-D point p is projected on the image at the intersection of the ray going through the optical center o and the image plane at a distance c. Note that $\tan(\alpha) = X/Z$

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. By the similarity of the triangles in the upper part of Fig. 9.26 the coordinates of the image are related to that of the object by *perspective projection*.

$$x' = -c\,\frac{X}{Z}\,, \quad y' = -c\,\frac{Y}{Z}\,. \tag{9.21}$$

The negative sign in (9.21) 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 Fig. 9.26 as virtually displacing the image plane in front of the optical center, which is the so-called *frontal pinhole camera model*. We define the camera frame with x' and y'-axes 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 depending on the arbitrary choice of image coordinates. Applying the change of sign to (9.21) and combining both coordinates into a vector yields

$$\boldsymbol{x}' = \begin{pmatrix} x' \\ y' \end{pmatrix} = \frac{c}{Z}\begin{pmatrix} X \\ Y \end{pmatrix}, \tag{9.22}$$

or equivalently in a homogeneous form

$$Z\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} c & 0 & 0 & 0 \\ 0 & c & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}, \tag{9.23}$$

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

9.3.4 Image Coordinates

In 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 in the upper left-hand 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. 9.27 the optical axis intersects the sensor at the *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 coordinate systems, units of which are specified in Table 9.7.

The transformation from pixel rows and columns $(\tilde{x}, \tilde{y})$ to a metric, sensor-centered image coordinate (x, y) with axis orientation as in Fig. 9.27 considers the position of PPS in pixels $(\tilde{x}_{\text{pps}}, \tilde{y}_{\text{pps}})$ and pixel size (e.g., in mm) along rows and columns $(\Delta_{\tilde{x}}, \Delta_{\tilde{y}})$

$$\begin{aligned} x &= \left[+\tilde{y} - \left(\frac{n_\text{c}}{2} - \frac{1}{2}\right)\right]\Delta_{\tilde{y}}\,, \\ y &= \left[-\tilde{x} + \left(\frac{n_\text{r}}{2} - \frac{1}{2}\right)\right]\Delta_{\tilde{x}}\,, \end{aligned} \tag{9.24}$$

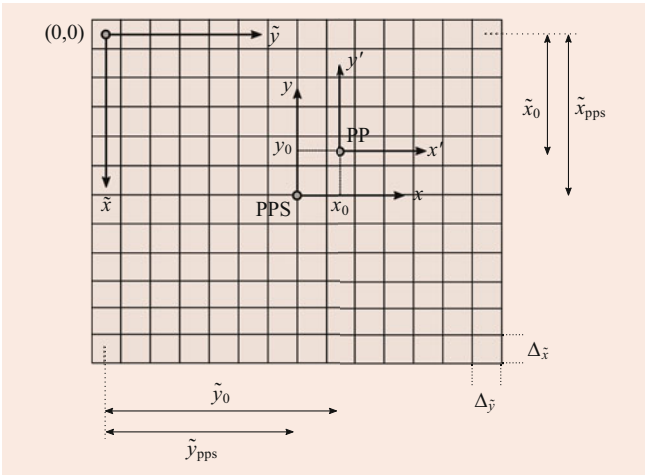

Fig. 9.27 Pixel $(\tilde{x}, \tilde{y})$, sensor-centered (x, y) and perspective-centered (x', y') image coordinates

Table 9.7 Definition of different image coordinate systems

Coordinates	Origin	Units	Usage
$(\tilde{x}, \tilde{y})$	Rows/cols counter	pixel	Computer vision
(x, y)	PPS	mm	Photogrammetry
(x', y')	PP	Unitless (= 1)	General

Table 9.8 Relations between image-coordinate systems

	$(\tilde{x}, \tilde{y})$	(x, y)	(x', y')
$(\tilde{x}, \tilde{y})$		(9.25)	(9.27)
(x, y)	(9.24)		(9.26)

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

$$\begin{pmatrix} \tilde{x} \\ \tilde{y} \\ 1 \end{pmatrix} = \begin{pmatrix} s_x & 0 & \tilde{x}_{\text{pps}} \\ 0 & s_y & \tilde{y}_{\text{pps}} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -y \\ x \\ 1 \end{pmatrix}, \tag{9.25}$$

with $\tilde{x}_{\text{pps}} = (n_r - 1)/2$, $\tilde{y}_{\text{pps}} = (n_c - 1)/2$ 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 autocollimation (PP). The orientation of the axis is arbitrary but in photogrammetry is usually defined as in Fig. 9.27. We also choose 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. 9.26 for x' and α. The transformation from so-called *reduced coordinates* (x', y') back to sensor-centered coordinates (x, y) in a homogeneous form is

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} c & 0 & x_0 \\ 0 & c & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix}, \tag{9.26}$$

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

$$\begin{pmatrix} \tilde{x} \\ \tilde{y} \\ 1 \end{pmatrix} = \begin{pmatrix} \tilde{c} & 0 & \tilde{x}_0 \\ 0 & \tilde{c} & \tilde{y}_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -y' \\ x' \\ 1 \end{pmatrix}, \tag{9.27}$$

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

Table 9.8 summarizes the transformations between the respective image-coordinate systems.

9.3.5 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 Sect. 9.3.2 that the mapping coordinates of a point $\boldsymbol{x}^m = (X^m, Y^m, Z^m)^\mathsf{T}$ relative to that of a camera $\boldsymbol{x}^c$ are related by the rigid body transformation (the inverse of (9.18))

$$\overline{\boldsymbol{x}}^c = \overline{\mathbf{T}}_c^{m-1} \overline{\boldsymbol{x}}^m, \tag{9.28}$$

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

Adopting the frontal camera model introduced in Sect. 9.3.3 for sensor-centered image coordinates (x, y) we rewrite (9.23) in vector notation

$$Z\,\overline{\boldsymbol{x}} = \begin{pmatrix} c & 0 & x_0 & 0 \\ 0 & c & y_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \overline{\boldsymbol{x}}^c \tag{9.29}$$

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 (9.29) into a product of two matrices while substituting for $\overline{\boldsymbol{x}}^c$ on the right-hand side with (9.28) we obtain the geometric model for *a basic camera*

$$\mu \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} c & 0 & x_0 \\ 0 & c & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{R}_c^m & \boldsymbol{x}_c^m \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} X^m \\ Y^m \\ Z^m \\ 1 \end{pmatrix}. \tag{9.30}$$

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

$$\mathbf{K} \doteq \begin{pmatrix} c & 0 & x_0 \\ 0 & c & y_0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \Pi_0 \doteq \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \tag{9.31}$$

we can rewrite the relation for a basic camera model (9.30) in a matrix form

$$\mu \overline{\boldsymbol{x}} = \mathbf{K} \Pi_0 \overline{\boldsymbol{x}}^c = \mathbf{K} \Pi_0 (\overline{\mathbf{T}}_c^m)^{-1} \overline{\boldsymbol{x}}^m = \Pi \overline{\boldsymbol{x}}^m, \tag{9.32}$$

when combining the 3×4 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-centered image coordinates (x'_d, y'_d) with radial symmetry as

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \left(1 + a_1 r^2 + a_2 r^4\right) \begin{pmatrix} x'_d \\ y'_d \end{pmatrix}, \tag{9.33}$$

where $r^2 = x'^2_d + y'^2_d$ is the square of the distance from the center and a_1, a_2 are the *distortion coefficients*. When needed, this simple radial model can be extended by additional coefficients, as in (9.70). Combining the relation of simple image distortion (9.33) together with basic camera

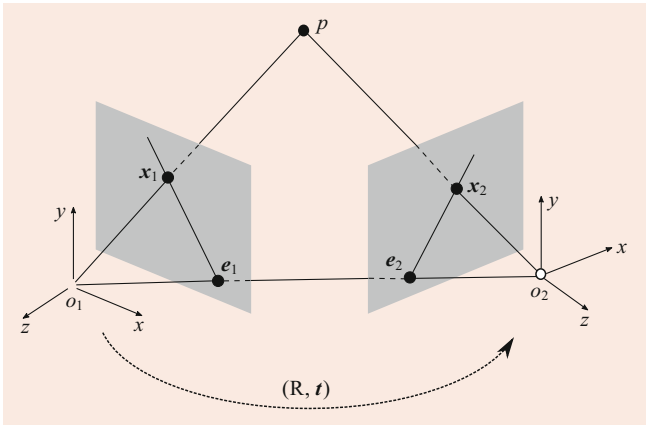

Fig. 9.28 Main steps of photogrammetry/computer-vision process in 3-D scene reconstructions based on 2-D imagery

projection model (9.32) we define the *realistic image formation model* that is applicable to many cameras employed in photogrammetry.

9.3.6 Scene from Two Views

The previously described image formation model (9.32) $\mu\overline{\boldsymbol{x}}' = \Pi\overline{\boldsymbol{x}}^m$ relates the object coordinates to image coordinates. Now we would like to perform the inverse and reconstruct 3-D object coordinates from images. As the scale (depth) μ is generally unknown due to 3-D $\longmapsto$ 2-D projection (note that μ varies per point and image), we need to employ at least two images of the same object with different camera poses that are known. Such a situation is depicted in Fig. 9.28; 3-D can be obtained 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.

Coplanarity Constraint

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

$$\mu_2 \boldsymbol{x}_2 = \mathbf{R}\mu_1 \boldsymbol{x}_1 + \boldsymbol{t} \ . \tag{9.34}$$

To eliminate the unknown depth μ we follow a couple of steps. First, we multiply both sides of the above equation from the left by a skew-symmetric matrix $[\boldsymbol{t}\times]$ containing vector $\boldsymbol{t}$. (This corresponds to cross product between two vectors.)

$$\mu_2 [\boldsymbol{t}\times]\boldsymbol{x}_2 = [\boldsymbol{t}\times]\mathbf{R}\mu_1 \boldsymbol{x}_1 + [\boldsymbol{t}\times]\boldsymbol{t} \ . \tag{9.35}$$

Due to orthogonality, the last term on the right-hand side is zero. Second, we multiply the last relation by $\boldsymbol{x}_2^\mathsf{T}$

$$\mu_2 \boldsymbol{x}_2^\mathsf{T}[\boldsymbol{t}\times]\boldsymbol{x}_2 = \boldsymbol{x}_2^\mathsf{T}[\boldsymbol{t}\times]\mathbf{R}\mu_1 \boldsymbol{x}_1 \ . \tag{9.36}$$

Since $[\boldsymbol{t}\times]\boldsymbol{x}_2$ is perpendicular to $\boldsymbol{x}_2$, and the inner product of the two perpendicular vectors is zero, the left-hand side $\boldsymbol{x}_2^\mathsf{T}[\boldsymbol{t}\times]\boldsymbol{x}_2 = 0$. Also, as $\mu_1 \neq 0$, we can write

$$\boldsymbol{x}_2^\mathsf{T}[\boldsymbol{t}\times]\mathbf{R}\,\boldsymbol{x}_1 = \boldsymbol{x}_2^\mathsf{T}\mathbf{E}\,\boldsymbol{x}_1 = 0 \ , \tag{9.37}$$

where $\mathbf{E} \doteq [\boldsymbol{t}\times]\mathbf{R}$ is called the *essential matrix*. The above relation is called an *epipolar constraint* as it conditions the three vectors $\boldsymbol{x}_2$, $\boldsymbol{t}$, and $\mathbf{R}\,\boldsymbol{x}_1$ to lie on a common plane, denoted as *epipolar plane*. Fig. 9.28 depicts also the two *epipoles* $\boldsymbol{e}_1, \boldsymbol{e}_2$ resulting from the intersection between a line $\mathbf{o}_1$–$\mathbf{o}_2$ and respective image planes. Connections $\boldsymbol{e}_1$–$\boldsymbol{x}_1$ and $\boldsymbol{e}_2$–$\boldsymbol{x}_2$ are called *epipolar lines*.

Essential Matrix Determination

To reconstruct $\mathbf{E}$ using only image observations, we briefly present the basic algorithm of [43] known also as the eight-point algorithm.

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

Second, we make use of the *Kronecker product* of two vectors,

$$\boldsymbol{a} \doteq \boldsymbol{x}_1 \otimes \boldsymbol{x}_2 = (x_1\boldsymbol{x}_2, x_2\boldsymbol{x}_2, x_3\boldsymbol{x}_2)^\mathsf{T} \ , \tag{9.38}$$

to express the epipolar constraint per one correspondence (point) as

$$\boldsymbol{a}^\mathsf{T} E^s = 0 \, . \tag{9.39}$$

Having a set of corresponding image points $(\boldsymbol{x}_1^i, \boldsymbol{x}_2^i)$, $i = 1, 2, \ldots, n$ we create n vectors $(\boldsymbol{a}^i)^\mathsf{T}$ and put them into a matrix $\boldsymbol{\chi} \doteq (\boldsymbol{a}^1; \boldsymbol{a}^2; \ldots; \boldsymbol{a}^n)$. With that we express the epipolar conditions for all correspondences in a system of linear equations

$$\boldsymbol{\chi} E^s = 0 \, . \tag{9.40}$$

The solution of this equation is unique if the rank of the matrix $\boldsymbol{\chi}$ is exactly 8 (a global scale factor cannot be determined). For this reason, we need $n \geq 8$ points. Note that expression (9.40) 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 eight-point algorithm, the

choice is made to minimize the least-squares error function of *misclosures* $|\boldsymbol{\chi} E^s|^2 \neq 0$, which is achieved by choosing E^s to be an eigenvector of $(\chi^\mathsf{T}\chi)$ that corresponds to its smallest singular value (eigenvalue) λ. Practically, this can be found by performing a singular value decomposition of $\boldsymbol{\chi} = \mathbf{U}_\chi \boldsymbol{\Sigma}_\chi \mathbf{V}_\chi^\mathsf{T}$; i.e., factoring χ into a product of diagonal matrix $\boldsymbol{\Sigma}_\chi$ containing the eigenvalues and orthogonal matrices ($\mathbf{U}^\mathsf{T}\mathbf{U} = \mathbf{I}$) $\mathbf{U}_\chi$ 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 $|\boldsymbol{\chi} E^s|^2$ in the least-squares sense, it is not guaranteed – due to the observation of unmodeled errors – that its structure belongs to the space of essential matrices. This space is characterized by $\mathbf{E} = \mathbf{U}\,\mathrm{diag}\{\sigma, \sigma, 0\}\mathbf{V}^\mathsf{T}$, where $\sigma = |\boldsymbol{t}|$. A common approach is, therefore, to reproject the estimated $\mathbf{E}$ to such a space. This is achieved by carrying the singular value decomposition of $\mathbf{E}$

$$\mathbf{E} = \mathbf{U}\,\mathrm{diag}\{\lambda_1, \lambda_2, \lambda_3\}\mathbf{V}^\mathsf{T}\,, \tag{9.41}$$

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 well be chosen as unity, which corresponds to *normalized essential space* where the two largest eigenvalues are set to 1.

Pose Reconstruction

Let us define a rotation matrix

$$\mathbf{R}_z(\pm\pi/2) = \begin{pmatrix} 0 & \mp 1 & 0 \\ \pm 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{9.42}$$

where the meaning of $\pm$ and $\mp$ signs will be explained later. Considering that (as for any rotation matrix) $\mathbf{R}_z\mathbf{R}_z^\mathsf{T} = \mathbf{I}$ together with the elements of singular value decomposition of $\mathbf{E}$, we can verify the correctness of the following relation

$$\begin{aligned} \mathbf{E} = [\boldsymbol{t}\times]\mathbf{R} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\mathsf{T} &= \mathbf{U}\mathbf{R}_z\boldsymbol{\Sigma}\mathbf{R}_z^\mathsf{T}\mathbf{V}^\mathsf{T} \\ &= \underbrace{\mathbf{U}\mathbf{R}_\mathbf{z}\boldsymbol{\Sigma}\mathbf{U}^\mathsf{T}}_{[\boldsymbol{t}\times]}\underbrace{\mathbf{U}\mathbf{R}_z^\mathsf{T}\mathbf{V}^\mathsf{T}}_{\mathbf{R}}\,, \end{aligned} \tag{9.43}$$

since $\mathbf{U}^\mathsf{T}\mathbf{U} = \mathbf{I}$ due to the orthonormality of $\mathbf{U}$ after reprojection of $\mathbf{E}$. Then the relative rotation follows directly from (9.43) as the product of three rotation matrices on the right, i.e.,

$$\mathbf{R} = \mathbf{U}\mathbf{R}_z^\mathsf{T}(\pm\pi/2)\mathbf{V}^\mathsf{T}\,, \tag{9.44}$$

and the relative translation (up to a scale) as

$$[\boldsymbol{t}\times] = \mathbf{U}\mathbf{R}_z^\mathsf{T}(\pm\pi/2)\boldsymbol{\Sigma}\mathbf{U}^\mathsf{T}\,, \tag{9.45}$$

where it can be proven that $\mathbf{U}\mathbf{R}_z^\mathsf{T}(\pm\pi/2)\boldsymbol{\Sigma}\mathbf{U}^\mathsf{T}$ 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 $(\mathbf{R}, \boldsymbol{t})$ from $\pm\mathbf{E}$. Three of these correspond to situations where the scene captured by one camera (first or second) or both cameras is situated behind the lens, which is physically not possible. These solutions can be eliminated by imposing the positive depth constraint. We should also mention that despite its simplicity, the eight-point algorithm is not without potential numerical weaknesses that may become apparent in a particular geometry and observation noise. However, as demonstrated by [44], these can be avoided by data preprocessing (translation and scaling).

Structure Reconstruction

With eight 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 the translation vector to unity is equivalent to 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 (9.34), which relates camera poses to n image correspondences

$$\mu_2^i \boldsymbol{x}_2^i = \mu_1^i \mathbf{R}\boldsymbol{x}_1^i + \xi\boldsymbol{t}, \quad i = 1, 2, \ldots, n\,. \tag{9.46}$$

Since $(\mathbf{R}, \boldsymbol{t})$ are known, this relation is linear and can, therefore, easily be solved once the unknown 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 $(\mathbf{R}, \boldsymbol{t})$, 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 $[\boldsymbol{x}_2\times]$ to obtain

$$\mu_1^i [\boldsymbol{x}_2^i\times]\mathbf{R}\boldsymbol{x}_1^i + \xi[\boldsymbol{x}_2^i\times]\boldsymbol{t} = 0\,. \tag{9.47}$$

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

$$\left([\boldsymbol{x}_2^i\times]\mathbf{R}\boldsymbol{x}_1^i, \quad [\boldsymbol{x}_2^i\times]\boldsymbol{t}\right)\begin{pmatrix}\mu_1^i \\ \xi\end{pmatrix} \doteq \mathbf{M}^i\boldsymbol{\mu}^i = 0\,. \tag{9.48}$$

To obtain a unique solution the matrix $\mathbf{M}$ needs to be of rank 1 or $[\boldsymbol{x}_2^i\times]\boldsymbol{t} \neq 0$. 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 $\overline{\boldsymbol{\mu}} =$

$(\mu_1^1, \mu_1^2, \ldots, \mu_1^n, \xi)^\mathsf{T}$ and a matrix $\mathbf{M}$ defined as

$$\mathbf{M} \doteq \begin{pmatrix} [\boldsymbol{x}_2^1\times]\mathbf{R}\boldsymbol{x}_1^1 & 0 & 0 & [\boldsymbol{x}_2^1\times]\boldsymbol{t} \\ 0 & [\boldsymbol{x}_2^2\times]\mathbf{R}\boldsymbol{x}_1^2 & 0 & [\boldsymbol{x}_2^2\times]\boldsymbol{t} \\ 0 & 0 & \vdots & 0 \\ 0 & 0 & [\boldsymbol{x}_2^n\times]\mathbf{R}\boldsymbol{x}_1^n & [\boldsymbol{x}_2^n\times]\boldsymbol{t} \end{pmatrix} . \tag{9.49}$$

The solution to the equation

$$\mathbf{M}\overline{\boldsymbol{\mu}} = 0 \tag{9.50}$$

determines all the unknowns in the vector $\overline{\boldsymbol{\mu}}$ 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 (9.50) can be found as the eigenvector of $\mathbf{M}^\mathsf{T}\mathbf{M}$ that corresponds to the smallest eigenvalue.

Global Scale

The global scale can be determined only by some exterior knowledge either on the camera motion, as discussed in Sect. 9.4.4, 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 three points in both frames to apply the seven-parameter similarity transformation. Finally, the problem of reconstruction can be formulated as an unconstrained optimization problem, where the minimization is searched with respect to all unknowns $\boldsymbol{x}_1^i, \mathbf{R}, \boldsymbol{t}, \bar{\mu}$. This is known in the literature as bundle adjustment, and its method will be further detailed in the next section, as well as in Sect. 9.4.3. The presented form, however, allows us to develop the needed approximations for its effective solution, which are based exclusively on image observations.

9.3.7 Scene from Multiple Views

Multiple-View Matrix

We now consider the existence of more than two views of the same object, which is rather a standard case. Without loss of generality, let us take the frame of the first camera as a reference frame for 3-D reconstruction. With m views/images at our disposition, from (9.32), we obtain the following projection matrices

$$\Pi_1 = (\mathbf{I}, \mathbf{0}), \Pi_2 = (\mathbf{R}_2, \boldsymbol{t}_2), \ldots, \Pi_m = (\mathbf{R}_m, \boldsymbol{t}_m) , \tag{9.51}$$

Considering at the moment only one point p and applying a similar development as for relations (9.46)–(9.48) we can derive the *multiple-view matrix* $\mathbf{M}_p$. We do so by inserting into two columns of $\mathbf{M}_p$ a coplanarity constraint (9.34) of view i between the first and i-th camera reference frame. Up to its depth this constraint is $\mu_1[\boldsymbol{x}_i\times]\mathbf{R}_i\boldsymbol{x}_1 + [\boldsymbol{x}_i\times]\boldsymbol{t}_i = 0$. In matrix form this is

$$\mathbf{M}_p \begin{pmatrix} \mu_1 \\ 1 \end{pmatrix} = 0 , \tag{9.52}$$

with $\mathbf{M}_p$ defined as

$$\mathbf{M}_p \doteq \begin{pmatrix} [\boldsymbol{x}_2\times]\mathbf{R}_2\boldsymbol{x}_1 & [\boldsymbol{x}_2\times]\boldsymbol{t}_2 \\ [\boldsymbol{x}_3\times]\mathbf{R}_3\boldsymbol{x}_1 & [\boldsymbol{x}_3\times]\boldsymbol{t}_3 \\ \vdots & \vdots \\ [\boldsymbol{x}_m\times]\mathbf{R}_m\boldsymbol{x}_1 & [\boldsymbol{x}_m\times]\boldsymbol{t}_m \end{pmatrix} . \tag{9.53}$$

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

$$\boldsymbol{x}_i^\mathsf{T}[\boldsymbol{t}_i\times]\mathbf{R}_i\boldsymbol{x}_1 = 0 . \tag{9.54}$$

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

Trilinear Constraint

In some situations, it may be useful to formulate one condition involving directly three views. Let us consider one point that is viewed by three cameras $1, i, j$. For this situation, we can write two separate coplanarity constraints, the second being transposed

$$\begin{aligned} \mu_1[\boldsymbol{x}_i\times]\mathbf{R}_i\boldsymbol{x}_1 &= -[\boldsymbol{x}_i\times]\boldsymbol{t}_i , \\ \boldsymbol{x}_1^\mathsf{T}\mathbf{R}_j^\mathsf{T}[\boldsymbol{x}_j\times]^\mathsf{T}\mu_1 &= -\boldsymbol{t}_j^\mathsf{T}[\boldsymbol{x}_j\times]^\mathsf{T} . \end{aligned} \tag{9.55}$$

Multiplying across the left and right-hand sides of both equations (9.55) and making them equal

$$-[\boldsymbol{x}_i\times]\mathbf{R}_i\boldsymbol{x}_1\boldsymbol{t}_j^\mathsf{T}[\boldsymbol{x}_j\times]^\mathsf{T} = -[\boldsymbol{x}_i\times]\boldsymbol{t}_i\boldsymbol{x}_1^\mathsf{T}\mathbf{R}_j^\mathsf{T}[\boldsymbol{x}_j\times]^\mathsf{T} , \tag{9.56}$$

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

$$[\boldsymbol{x}_i\times]\left(\boldsymbol{t}_i\boldsymbol{x}_1^\mathsf{T}\mathbf{R}_j^\mathsf{T} - \mathbf{R}_i\boldsymbol{x}_1\boldsymbol{t}_j^\mathsf{T}\right)[\boldsymbol{x}_j\times] = 0 . \tag{9.57}$$

The trilinear constraint implies a bilinear constraint (9.54), except for a special case in which $[\boldsymbol{x}_j\times]\boldsymbol{t}_j = [\boldsymbol{x}_j\times]\mathbf{R}_j\boldsymbol{x}_i = 0$ 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 the 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. (This may be the case, for instance, in a 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.) When they still satisfy the trilinear constraint, 3-D coordinates of this point can be reconstructed. It should also be 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).

Processing Strategies

The processing strategies for handling multiple views vary in terms of image geometry, scene texture (goodness of feature detection, matching, and filtering), camera calibration, data noise, and experience. We therefore present only the main concepts while leaving out the details of their combination into to a particular implementation. These are schematically depicted in Fig. 9.29.

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 Fig. 9.29 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 eight-point algorithm only once for some initial pairs of view and perform global optimization on the rest, as shown in the bottom part of Fig. 9.29. This method is more suitable for good image geometry, large overlap, and (pre)calibrated cameras. It involves constructing a relation containing the multiple-view matrix $\mathbf{P}_i$, similar to that of (9.53) but involving m images $\boldsymbol{x}_1^j, \boldsymbol{x}_2^j, \ldots, \boldsymbol{x}_1^m$ of n points p^j, $j = 1, 2, \ldots n$ from which we would like to estimate the unknown projections $\Pi_i(\mathbf{R}_i^s, \boldsymbol{t}_i)^\mathsf{T}$, $i = 2, 3, \ldots, m$

$$\mathbf{P}_i \begin{pmatrix} \mathbf{R}_i^s \\ \boldsymbol{t}_i \end{pmatrix} \begin{pmatrix} [\boldsymbol{x}_1^1\times]^\mathsf{T} \otimes [\boldsymbol{x}_i^1\times] & \lambda^1[\boldsymbol{x}_i^1\times] \\ [\boldsymbol{x}_1^2\times]^\mathsf{T} \otimes [\boldsymbol{x}_i^2\times] & \lambda^2[\boldsymbol{x}_i^2\times] \\ \vdots & \vdots \\ [\boldsymbol{x}_1^n\times]^\mathsf{T} \otimes [\boldsymbol{x}_i^n\times] & \lambda^n[\boldsymbol{x}_i^n\times] \end{pmatrix} \begin{pmatrix} \mathbf{R}_i^s \\ \boldsymbol{t}_i \end{pmatrix} = 0\,, \tag{9.58}$$

where $\otimes$ is the *Kronecker product* between matrices (similarly to that of (9.38) but with the elements of the left matrix stacked into a vector), and 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 $\Pi_i = (\mathbf{R}_i, \boldsymbol{t}_i)$ 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 $\mathbf{R}_i$ needs to be reprojected to the rotational space $SO(3)$ and the vector $\boldsymbol{t}_i$ rescaled. Assuming the pose for the second view is found by the eight-point algorithm, the scalars λ^j can be determined from the first row of (9.58) involving $\lambda^j[\boldsymbol{x}_2^j\times]\boldsymbol{t}_2 = -[\boldsymbol{x}_2^j\times]\mathbf{R}_2\boldsymbol{x}_1^j$. These initial values of λ^j can then be used for the recovery of Π_i, $i = 3, 4, \ldots, m$. Since $\boldsymbol{t}_2$ is recovered from the eight-point algorithm up to a scale factor, the other views are recovered from that up to a global single scale.

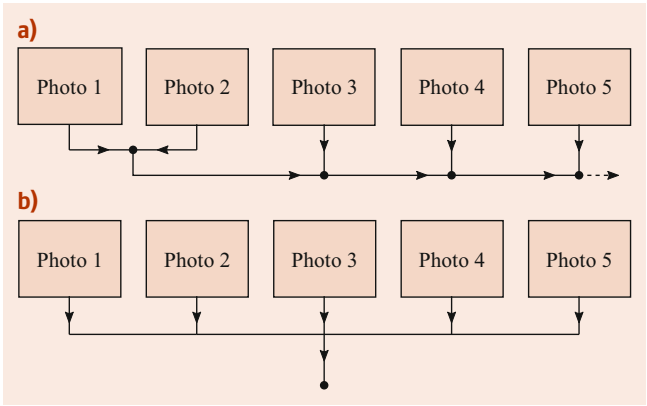

Fig. 9.29 Image reconstruction strategies: incremental (**a**) versus global (**b**)

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 also accommodates other inputs, is presented in Sect. 9.4 on sensor fusion. Bundle adjustment received its name after application of the ray-tracing collinearity condition (9.32) $\mu\overline{\boldsymbol{x}'} = \Pi\overline{\boldsymbol{x}}^m$ on a "bundle of rays" connecting object points with its projection on the image in combination with a particular sensor model (9.33). Nevertheless, this optimization approach requires linearization and 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.

9.3.8 Feature Matching

The term image matching stands for the mostly automatic reference between regions or pixels of two or more im-

ages 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 that are geometrically stable under different transformations and have a 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 results 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.

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 [45, 46]. 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 [47]. 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 [48] 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 [49]. Meanwhile, the scale-invariant feature transform (SIFT) keypoint detector [50] 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 in another can still 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) [50]. Thus, it has become especially popular in close-range applications, where matching is frequently aggravated due to the appearance of larger perspective distortions.

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 feature's 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 keypoint location [50]. This approach transforms the image data into a scale and rotation-invariant representation. A pair of keypoints 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 [51].

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)

$$\rho(r,c) = \frac{\left(\begin{array}{c} \sum_{i=1}^{m}\sum_{j=1}^{n}[g_{\mathrm{L}}(i,j) - \bar{g}_{\mathrm{L}}] \\ \cdot\,[g_{\mathrm{R}}(r+i,c+j) - \bar{g}_{\mathrm{R}}] \end{array} \right)}{\sqrt{\begin{array}{l} \sum_{i=1}^{m}\sum_{j=1}^{n}[g_{\mathrm{L}}(i,j) - \bar{g}_{\mathrm{L}}]^2 \\ \cdot \sum_{i=1}^{m}\sum_{j=1}^{n}[g_{\mathrm{R}}(r+i,r+j) - \bar{g}_{\mathrm{R}}]^2 \end{array}}}, \tag{9.59}$$

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

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 the epipolar geometry of the respective image pair are frequently used. As an example, algorithms based on random sample consensus (RANSAC) [52] 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 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 [53], or, alternatively, with (9.43) when setting the translation vector to unity.
3. This relative orientation defines for each image point $j = 1, 2, \ldots, n$ in the left image the corresponding epipolar line in the right image (9.37), $\boldsymbol{x}_2^j \mathbf{E} \boldsymbol{x}_1^j = 0$. 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. 9.30.
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 corresponding image points.

These correspondences can be directly used as tie-points during optimization 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 NCC as defined by (9.59). 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 directions are parabolas.

$$f(r,c) = a_0 + a_1 r + a_2 c + a_3 r^2 + a_4 rc + a_5 c^2. \quad (9.60)$$

$$\begin{pmatrix} \rho_0(-1,-1) & \rho_1(-1,0) & \rho_2(-1,1) \\ \rho_3(0,-1) & \rho_4(0,0) & \rho_5(0,1) \\ \rho_6(1,-1) & \rho_7(1,0) & \rho_8(1,1) \end{pmatrix}, \quad (9.61)$$

which represents the NCC coefficients for a $\times$ local neighborhood centered at position $(0,0)$ of the maximum value p_4. The NCC coefficients $\boldsymbol{l} = (\rho_0, \rho_1, \ldots, \rho_8)^\mathsf{T}$ computed by (9.59) for the different positions (r_i, c_i) are then used as observations within the Gauss–Markov model $\mathbf{A}\boldsymbol{x} - \boldsymbol{l} = \boldsymbol{v}$. The parameters of the polynomial (9.60) can then be estimated with

$$\mathbf{A} = \begin{pmatrix} 1 & r_0 & c_0 & r_0^2 & r_0 c_0 & c_0^2 \\ 1 & r_1 & c_1 & r_1^2 & r_1 c_1 & c_1^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & r_8 & c_8 & r_8^2 & r_8 c_8 & c_8^2 \end{pmatrix}. \quad (9.62)$$

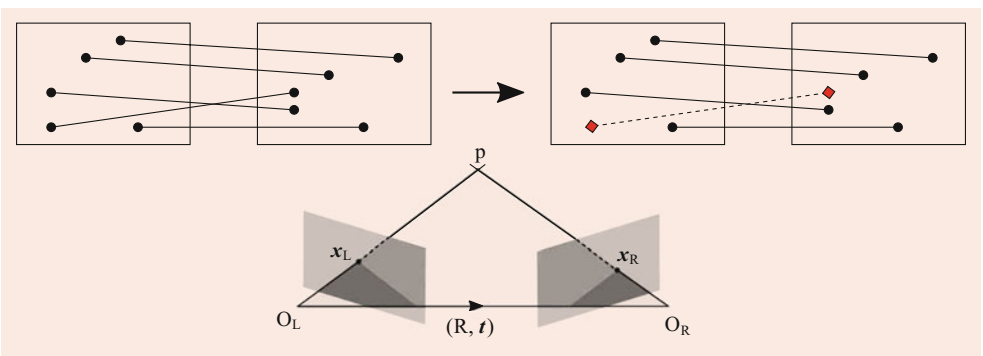

Fig. 9.30 Filtering of wrongly assigned correspondences on a randomly selected subset of points through the epipolar constraint

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

$$\begin{pmatrix} 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 0 & 1 & 0 & 0 \\ 1 & -1 & 1 & 1 & -1 & 1 \\ 1 & 0 & -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{pmatrix} - \begin{pmatrix} \rho_0 \\ \rho_1 \\ \rho_2 \\ \rho_3 \\ \rho_4 \\ \rho_5 \\ \rho_6 \\ \rho_7 \\ \rho_8 \end{pmatrix} = V \ . \quad (9.63)$$

The standard solution

$$\boldsymbol{x} = (\mathbf{A}^\mathsf{T}\mathbf{A})^{-1}\mathbf{A}^\mathsf{T}\boldsymbol{l} = [\mathbf{A}(\mathbf{A}^\mathsf{T}\mathbf{A})^{-1}]^\mathsf{T}\boldsymbol{l} = \boldsymbol{t}'\boldsymbol{l}$$

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

$$\frac{\partial f}{\partial r} = a_1 + 2a_3\Delta r + a_4\Delta c \cong 0 \ , \quad (9.64)$$

$$\frac{\partial f}{\partial c} = a_2 + a_4\Delta r + a_5\Delta c \cong 0 \ . \quad (9.65)$$

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

$$\Delta r = \frac{a_2a_4 - 2a_1a_5}{4a_3a_5 - a_4^2} \ , \quad (9.66)$$

$$\Delta c = \frac{a_1a_4 - 2a_2a_3}{4a_3a_5 - a_4^2} \ . \quad (9.67)$$

An example input for the computation of subpixel refinement can be found in Fig. 9.31, which depicts a correlation mask in the left-hand image and a search mask in the right-hand 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-hand feature point give the matrix

$$\begin{pmatrix} 0.9300 & 0.9088 & 0.8622 \\ 0.9862 & 0.9922 & 0.9664 \\ 0.9281 & 0.9646 & 0.9696 \end{pmatrix} . \quad (9.68)$$

If these values are used as input for (9.63) and (9.66), 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 keypoint extraction [50]. In this application, pixel coordinates x, y and scale s define a 3-D scale space function $D(x, y, s)$. Thus, subpixel and subscale coordinates $x = (x; y; s)$ of a feature are found by interpolation with a 3-D quadratic function that 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 [48] can be applied. This approach estimates the geometric and radiometric transformations between corresponding patches g_L and g_R from the left-hand and right-hand images, respectively, using the observation equation

$$g_L(r, c) + v = h_1 + h_2 g_R[(a_0 + a_1 r + a_2 c), (b_0 + b_1 r + b_2 c)] \ . \quad (9.69)$$

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.

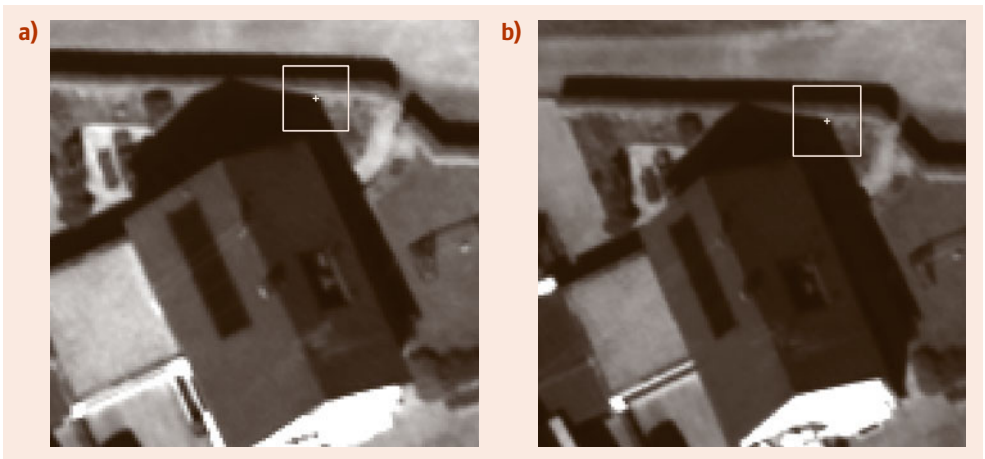

Fig. 9.31 Feature point with correlation mask (**a**) and search mask (**b**)

Table 9.9 Participants of the EuroSDR-Project Benchmark on Image Matching, 2014

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

Equation 9.69 equals the Gauss–Markov model, which minimizes the squared sum of errors of all the observations $v^\mathsf{T}v \rightarrow \min$. Gray-value differences between the corresponding image patches are used for a typical window size of 15×15 pixel, resulting in a matching precision of 0.1–0.01 pixels.

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 at dense, pixel-wise matches. By these means, 3-D point clouds and digital surface models (DSM) are generated at a resolution that 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 as a global optimization problem [54]. In a way, these global algorithms allow 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. 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 semiglobal matching [55], which evaluates a cumulative cost function from the scanlines in the four cardinal and four ordinal directions, east, northeast, north etc., Though the algorithm operates in two dimensions, it is still fast, because it substitutes the 2-D computation by 8 1-D computations. Especially when it is combined with sophisticated aggregation strategies it can produce accurate results very efficiently [54].

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) [56]. This benchmark tested the DSM programs listed in Table 9.9.

In addition to the generation of 2.5-D models like DSM and DTM, the extraction of real 3-D structures especially in dense and complex urban environments is becoming increasingly 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 at semantic interpretation of 3-D data.

9.4 Sensor Fusion

9.4.1 Principle

In Sect. 9.3, we described how to reconstruct a 3-D scene using structure-from-motion techniques. Such methods rely solely on image observations to determine the relative orientation between images at an arbitrary scale (e.g., 1). Other observations need 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 such a way that allows

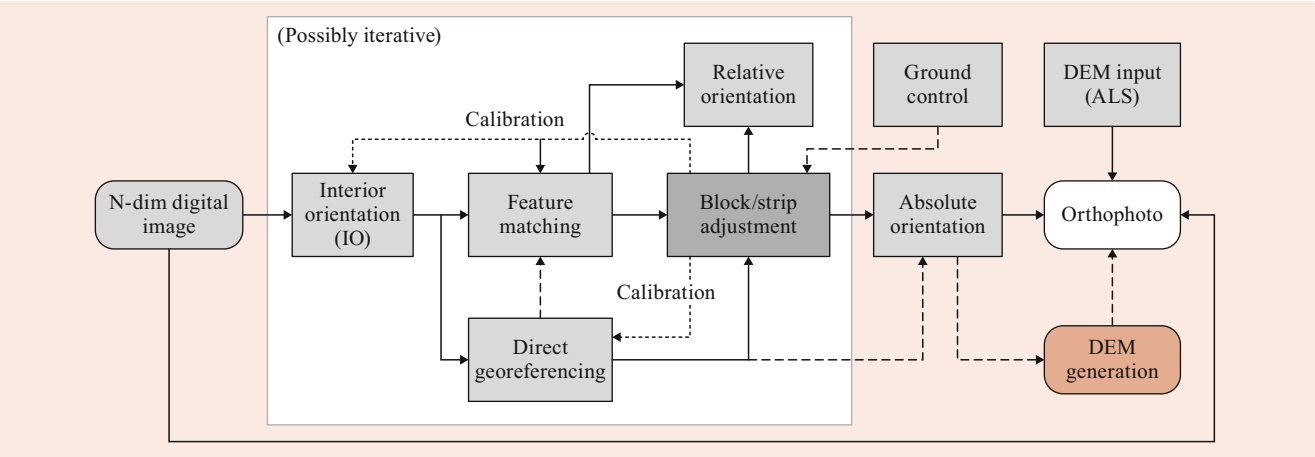

Fig. 9.32 General overview of sensor fusion, with the optimization step denoted as block/strip adjustment, also known as assisted aerotriangulation (AAT)

an optimal recovery of all parameters involved, including those related to the unknown parameters of optical sensors for the purpose of mapping. This, generally, requires determining:

1. The sensors' interior orientation (IO) parameters. In the 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 (in the case of lidars this may be rangefinder bias or misalignment between the laser-beam reflecting mirror and its angular encoder).
2. The sensors' absolute exterior orientation (EO) that, similarly to relative orientation, geometrically connects the sensor data among them, plus with respect to the real world (Sect. 9.2) (this is also referred to as pose when limited to position and attitude).
3. Auxiliary system parameters that geometrically relate 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. 9.32, the process of sensor orientation may take different paths (dashed versus full lines):

- The sensor data can be oriented directly using navigation technology (Sect. 9.2).
- The sensor is oriented indirectly by identifying corresponding features across overlapping parts of data and connecting them with external references on the ground (Sect. 9.3.7). This is achieved by a procedure that will be referred to as bundle block (or strip) adjustment. In the context of airborne mapping with passive imagery, this approach is called aerial triangulation (AT).

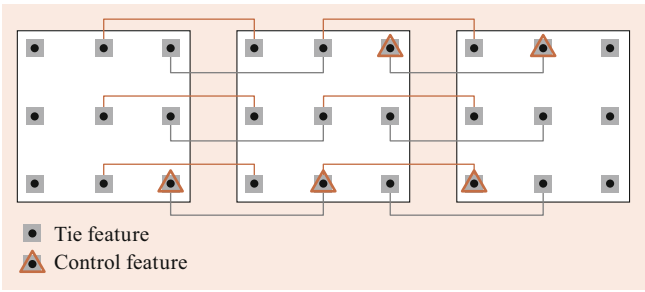

Fig. 9.33 Feature (point) conditioning on overlapping imagery

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

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. 9.33). 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 the same coordinates as in the object coordinate system. This approach is applicable to passive (Sect. 9.1.1) as well as active (Sect. 9.1.3) 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 part of Fig. 9.34.

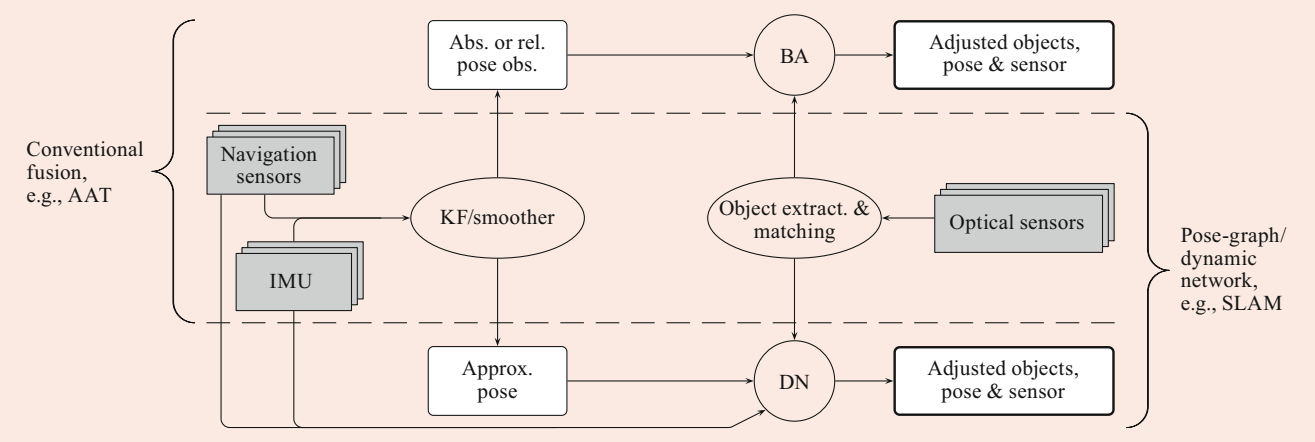

Fig. 9.34 Comparison of traditional and modern methods for fusing data from optical and navigation sensors in mapping

Table 9.10 Example of parameters for sensor orientation and calibration

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

If satellite positioning 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, the character of which cannot be correctly modeled within the network of this type. In such a situation it 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 part of Fig. 9.34. This approach is rigorous but requires some special care when introducing the differential equations relating the inertial observation to poses. This method proposed by [57] under the term dynamic network bears a number of advantages as well as challenges. With a number of simplifications that are not appropriate for airborne mapping, this approach is also employed in robotic indoor SLAM where it is referred to as pose-graph estimation [59]. Its extended modeling, which is applicable to large-scale mapping project, is described in [60].

9.4.2 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 Table 9.10. 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), $\boldsymbol{p}_n^m$, with $n \in \{1; \ldots; N\}$
- (optionally) basic interior orientation parameters, i.e., the camera constant c, the principal point (x_0, y_0)
- (optionally) additional interior orientation parameters represented either by physical or the replacement models (Sect. 9.4.3)
- 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; $\Gamma_{b,j}^m = [\boldsymbol{x}_b^m(j), \mathbf{R}_b^m(j)]$, with $j \in \{1; \ldots; J\}$
- (optionally) the body to camera lever-arm and bore sight $\boldsymbol{x}_{bs}^b$ and $\mathbf{R}_s^b$
- (optionally in dynamic network/SLAM or in AAT if GNSS position is used instead of GNSS/INS) GNSS antenna lever-arm $\boldsymbol{x}_{ba}^b$, or $\boldsymbol{x}_{sa}^s$, respectively
- (optionally in dynamic network/SLAM) INS systematic errors, e.g., random, yet time-constant 3-D bias vectors for the gyroscopes $\boldsymbol{b}_\omega$ and accelerometers $\boldsymbol{b}_f$.

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

9.4.3 Optical Distortion Models

Optical distortions directly influence the metric quality of the image and, therefore, have to be considered. As introduced in Sect. 9.1 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 modeled (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 relate 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.

Sensor Physical Models

In many frame/line cameras, symmetric lens distortion has the most relevant influence on 3-D object point reconstruction. Relation (9.33) introduced a basic distortion model of perspective-centered image coordinates. A more general model is the Conrady–Brown distortion correction [61] 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

$$\begin{aligned} x_d &= x\left(k_1 r^2 + k_2 r^4 + k_3 r^6\right) + p_1\left(r^2 + 2x^2\right) + 2p_2 xy \,, \\ y_d &= y\left(k_1 r^2 + k_2 r^4 + k_3 r^6\right) + p_2\left(r^2 + 2y^2\right) + 2p_1 xy \,, \end{aligned} \tag{9.70}$$

with $r^2 = x^2 + y^2$. Affinity as well as nonorthogonality effects on image coordinates maybe added to (9.70), and the size of a particular calibration set may be even larger depending on the system and the type of calibration [62]. 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 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. 9.35, 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.

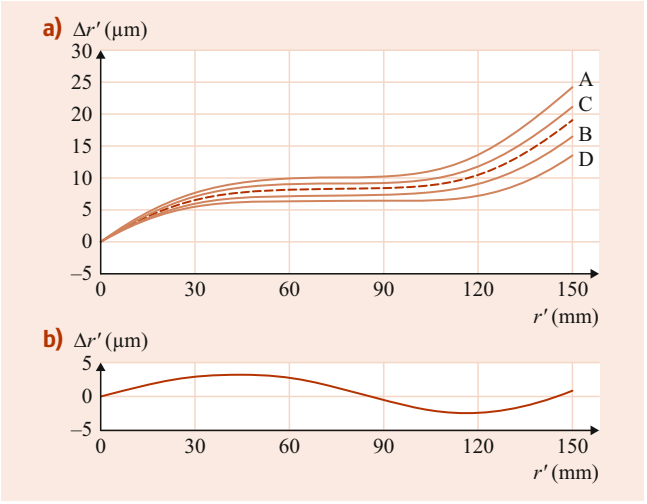

Fig. 9.35 (**a**) Radial distortion curves (A–D) and their mean value (*dashed line*). (**b**) Mean radial distortion after balancing, after [2]

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, possibly being 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:

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

9.4.4 Observation Models

Image Observations

Frame cameras

Here we combine the basic camera model (9.32) that gives the undistorted image coordinates of a mapped feature and the geometrical relations between the optical and navigation data (Fig. 9.23 and Sect. 9.2.5). Let $\boldsymbol{p}_n^m$ be the 3-D coordi-

nates of the n-th tie-point expressed in the mapping frame. Considering the poses of the b-frame associated with the j-image $\Gamma^m_{b,j} = [\boldsymbol{x}^m_b(j), \mathbf{R}^m_b(j)]$, through the camera bore sight $\mathbf{R}^b_s$ and lever-arm $\boldsymbol{x}^b_{bs}$, 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

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \frac{1}{\mu} \begin{pmatrix} c & 0 & x_0 \\ 0 & c & y_0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \left[\left(\mathbf{R}^m_b(j)\mathbf{R}^b_s\right)^\mathsf{T} \left(\boldsymbol{p}^m_n - \boldsymbol{x}^m_b(j)\right) - \left(\mathbf{R}^b_s\right)^\mathsf{T} \boldsymbol{x}^b_{bs} \right]. \tag{9.71}$$

The scale factor μ can be eliminated by rearranging the equation system (9.71) so that the image coordinates are separated on the left-hand side and then dividing the first two relations. Then, the distorted image coordinates $(x_\mathrm{d}, y_\mathrm{d})$ can be determined, e.g., as in (9.70). Finally, with $\boldsymbol{l}_{n,i}$ 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

$$\boldsymbol{l}_{n,i} + \boldsymbol{v}_{n,i} = \begin{pmatrix} x_\mathrm{d} \\ y_\mathrm{d} \end{pmatrix}, \tag{9.72}$$

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

Line cameras

The collinearity condition expressed for frame cameras (9.71) with the column (x) and row (y) image coordinates can be adapted to line cameras by omitting the y or row pixel coordinates

$$\begin{pmatrix} x \\ 0 \\ 1 \end{pmatrix} = \frac{1}{\mu} \begin{pmatrix} c & 0 & x_0 \\ 0 & c & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \left[\left(\mathbf{R}^m_b(j)\mathbf{R}^b_s\right)^\mathsf{T} \left(\boldsymbol{p}^m_n - \boldsymbol{x}^m_b(j)\right) - \left(\mathbf{R}^b_s\right)^\mathsf{T} \boldsymbol{x}^b_{bs} \right]. \tag{9.73}$$

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 $\mathbf{K}$ (9.31), the undistorted collinearity model reads

$$\begin{pmatrix} x_k \\ 0 \\ 1 \end{pmatrix} = \frac{1}{\mu} \mathbf{K} \left[\left(\mathbf{R}^m_b(j)\mathbf{R}^b_p\mathbf{R}^p_{s,k}\right)^\mathsf{T} \left(\boldsymbol{p}^m_n - \boldsymbol{x}^m_b(j)\right) - \left(\mathbf{R}^b_p\right)^\mathsf{T} \boldsymbol{x}^b_p + \boldsymbol{x}^p_{bs,k} \right], \tag{9.74}$$

where p is the common camera-platform reference frame, $\boldsymbol{x}^p_{bs,k}$ is the lever-arm of line k to p-frame origin, and $\mathbf{R}^p_{s,k}$ is the rotation from the k-camera to the platform frame. With $\boldsymbol{l}^k_{n,i}$ 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

$$\boldsymbol{l}_{m,n,i} + \boldsymbol{v}_{m,n,i} = \boldsymbol{x}_{d,m}\,. \tag{9.75}$$

The relation between distorted x_k and undistorted image coordinates $x_{d,k}$ depends on the optical model described in Sect. 9.4.3, e.g., the first line of (9.70).

Ground Control

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

$$\boldsymbol{l}_n + \boldsymbol{v}_n = \boldsymbol{p}^m_n\,. \tag{9.76}$$

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 latter 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 the m-frame. Referring to (9.14) and considering the positions $\boldsymbol{x}^m_{a/b}$ with respect to m together with the lever-arm $\boldsymbol{x}^{a/b}_s$ and attitude $\mathbf{R}^m_b$, the observation model for sensor position reads

$$\boldsymbol{l}_{p,j} + \boldsymbol{v}_{p,j} = \boldsymbol{x}^m_{a/b}(j) + \mathbf{R}^m_b(j)\boldsymbol{x}^{a/b}_s\,. \tag{9.77}$$

When the chosen mapping frame is a projection and the sensor-fusion model is derived for a Cartesian frame, the position observation should be corrected in height. As described in [30], the size of such a correction depends on the absolute terrain height, the flying altitude above it, and the value of the projection scale at the perspective center.

Velocity

Velocity observations are needed for sensors like radars. They can be also useful for estimating a time-stamping offset between optical and navigation data [63]. GNSS or GNSS/INS provide velocity observation of the antenna a or body b, either in e or l frames, respectively. Similarly to position, the velocity vector can be transformed to a mapping frame. The velocity observation model is then

$$\boldsymbol{l}_v + \boldsymbol{v}_v = \boldsymbol{v}^m_{a/b}(j) + \mathbf{R}^m_b(j)\boldsymbol{\Omega}^b_{mb}(j)\boldsymbol{x}^{a/b}_s\,, \tag{9.78}$$

where $\boldsymbol{\Omega}^b_{mb}(j)$ is the skew-symmetric matrix of angular velocity vector $(\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3) = (\Omega_{3,2} = -\Omega_{2,3}, \Omega_{1,3} = -\Omega_{3,1}, \Omega_{2,1} = -\Omega_{1,2}))$ between the m and b-frames expressed in the b-frame.

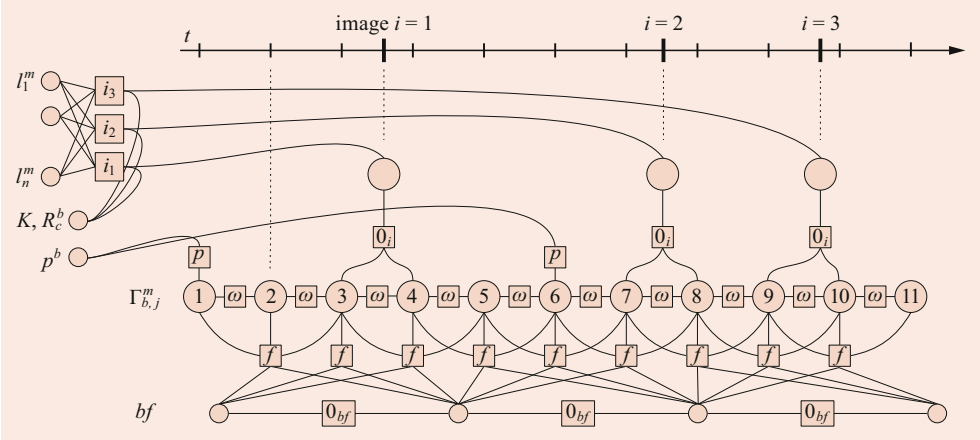

Fig. 9.36 A simplified instance of a dynamic network formulated as a factor graph

Attitude

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

$$\boldsymbol{l}_R + \boldsymbol{v}_R \equiv \mathbf{R}_m^b \, \mathbf{R}_v = \mathbf{R}_s^b \, \mathbf{R}_m^s \,, \tag{9.79}$$

with $\mathbf{R}_s^b$ being the bore sight and $\mathbf{R}_m^s$ the mapping-to-sensor frame rotation. However, INS/GNSS processing usually delivers $\mathbf{R}_l^b$, where the orientation of the local-level frame l with respect to the m-frame changes with the change of position. In such a case, the attitude observation equations need to be modified to

$$\mathbf{R}_l^b \, \mathbf{R}_v = \mathbf{R}_s^b \, \mathbf{R}_m^s \, \mathbf{R}_l^m \,, \tag{9.80}$$

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

$$\mathbf{R}_l^{m \equiv l_0} = \mathbf{R}_e^l(\phi_0, \lambda_0) \, \mathbf{R}_l^e(\phi_t, \lambda_t) \,, \tag{9.81}$$

where $\mathbf{R}_l^e = (\mathbf{R}_e^l)^\intercal$ is defined by (9.11).

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 for. Considering the east-north-up axis convention usually used in projections, the attitude observation for the projection reads

$$\mathbf{R}_l^{m \equiv p} = \mathbf{R}_{\mathrm{NED}}^{\mathrm{ENU}} \, \mathbf{R}_3(\gamma_{\mathrm{PC}}) \,, \tag{9.82}$$

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 the national datum [31].

Angular Velocity and Specific Forces

As outlined in Sect. 9.4.1, some method of robotic'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 in [60]. As schematically depicted in the bottom part of Fig. 9.36, ω and f constrain the unknown poses $\Gamma_{b,j}^m$ via differential relations that are approximated by first and second-order finite differences. Also shown schematically in Fig. 9.36 are the connections between other previously mentioned observation models (represented by boxes) with the unknown parameters (represented by circles): p for GNSS positions, l_n^m for image observations, and 0 for so-called zero observations. The latter relates some parameters by known functional relationships 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. 9.36 for the sake of clarity.

9.4.5 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 $\boldsymbol{v}$ of all observations are randomly distributed, this is achieved by solving a nonlinear weighted least-squares problem. Gathering all the terms on one side of each observation

equation will result in the following condition

$$g(\boldsymbol{l} + \boldsymbol{v}, \boldsymbol{x}) = 0\,, \tag{9.83}$$

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

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

$$g(\boldsymbol{l}, \boldsymbol{x}_i) + \left(\frac{\partial g}{\partial \boldsymbol{l}}\right)_{|(\boldsymbol{l}, \boldsymbol{x}_i)} v + \left(\frac{\partial g}{\partial \boldsymbol{x}}\right)_{|(\boldsymbol{l}, \boldsymbol{x}_i)} \Delta\, \boldsymbol{x}_i = 0 \tag{9.84}$$

  or

$$-\,\boldsymbol{g}_i - \mathbf{B}_i\, \boldsymbol{v} + \mathbf{A}_i\, \Delta\, \boldsymbol{x}_i = 0\,, \tag{9.85}$$

  where $\boldsymbol{g}_i = -g(\boldsymbol{l}, \boldsymbol{x}_i)$, $\mathbf{B}_i = -(\partial g/\partial \boldsymbol{l})_{|(\boldsymbol{l}, \boldsymbol{x}_i)}$, and $\mathbf{A}_i = (\partial g/\partial \boldsymbol{x})_{|(\boldsymbol{l}, \boldsymbol{x}_i)}$.
  (In the case of $\boldsymbol{v} = -\boldsymbol{l} + g(\cdot)$, this derivative is trivial.)
- Second, the nonlinear model (9.83) is solved by iterating the solutions of the linear model (9.85) to convergence. The corrections to the parameters Δx_{i+1} are obtained by the best linear unbiased estimation (BLUE). As for the parameters, the BLUE estimation is the solution of the so-called normal equation [64]

$$\Delta \boldsymbol{x}_i = \left(\mathbf{A}^{\mathsf{T}}(\mathbf{B}\mathbf{C}_{vv}\mathbf{B}^{\mathsf{T}})^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathsf{T}}(\mathbf{B}\,\mathbf{C}_{\ell\ell}\mathbf{B}^{\mathsf{T}})^{-1}\boldsymbol{g}_i\,, \tag{9.86}$$

  while the corrections to observations are obtained as

$$\boldsymbol{v}_i = \mathbf{C}_{vv}\mathbf{B}^{\mathsf{T}}(\mathbf{B}\mathbf{C}_{\ell\ell}\mathbf{B}^{\mathsf{T}})^{-1}(\mathbf{A}\boldsymbol{x} - \boldsymbol{g}_i)\,. \tag{9.87}$$

After each iteration, the set of parameters is rectified by the estimated correction $\Delta \boldsymbol{x}_i$ as $\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \Delta\, \boldsymbol{x}_i$, and the observations are updated according to $\boldsymbol{l}_{i+1} = \boldsymbol{l}_i + \boldsymbol{v}_i$. The linearization of (9.84) 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 $\Delta\, \boldsymbol{x}_i$ are no longer significant (i.e., $\Delta\, \boldsymbol{x}_i \approx 0$). After convergence, the last iteration step is repeated with the original observations ($\boldsymbol{l}_{i=0}$). The respective covariances characterizing the accuracy of parameters and measurements are estimated in parallel at each step by the relations presented in [64].

The quality of the estimation may be judged according to the analyses of residuals and global a-posteriori estimation of the variance. The latter is evaluated as $\widehat{\sigma}_0{}^2 = (\boldsymbol{v}^{\mathsf{T}}\, \mathbf{C}_{\ell\ell}{}^{-1}\, \boldsymbol{v})/(n-u)$, where n is the number of observations and u the number of parameters. Special situations may lead to some simplification of the general model (9.84) and its solution (9.86). Detailed information on this subject is presented in [65] and [64].

The general formulation of the sensor fusion may be very large, leading to hundreds of thousands unknowns (or even millions of unknowns for dynamic networks) but is inherently sparse and can be solved efficiently exploiting state-of-the-art least-squares solvers [66] based on very efficient sparse linear algebra routines [67]. (Reducing the number of unknowns in dynamic network is possible by preintegrating a certain number of IMU observations [68].)

9.4.6 Adopted Approaches

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 aerotriangulation 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 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 remains in use for small mapping projects, such as those performed by small drones without RTK capacity.

GNSSs are included for quality control or for calibration purposes on 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 (Table 9.11). 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 autopiloted drones (for the purpose of navigation, guidance, and control capacity on automated missions). 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 than 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 pur-

Table 9.11 Comparison of main orientation approaches

Aerial Observations	Advantages	Disadvantages
None	• Independent of airborne satellite signal • Simple processing chain • Independent of navigation quality • Independent of synchronization errors	• Impractical 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	• Weak 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 & multisensor systems • ≈20% side-overlap OK • Automation, no GCPs	• Lower redundancy in corridors • Attitude accuracy dependent

Table 9.12 Indicative frequency of sensor deployment and orientation method used for frame cameras across different platforms: (−) = never, (+) = rare, (++) = sometimes, (+++) = common

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

poses. Similarly to position-assisted AT, the use of full aerial control results in lower correlation between EO/IO parameters. Table 9.12 indicates common orientation approaches across different platforms. Generally, it is acknowledged that object–space accuracy is two to four times better when using an integrated rather than a direct sensor orientation approach.

Line Sensors

Theoretically, the orientation of line sensor data can also be determined indirectly, in a manner similar to frame imagery [69]. However, this approach is rarely used in practice because the computational effort is large, and the resulting mapping accuracy is lower without the support of attitude and position sensors [70]. 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 a 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. 9.37. The onboard GNSS/INS measurements are recorded for postprocessing (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 during postprocessing using the best available calibration parameters and the improved EO parameters from the postprocessed trajectory. The distortion in the original imagery, caused by the 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 raw images for the (best possible) orthophoto production.

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 the use of a postprocessed trajectory [70]. Also, the accuracy of object points was two to three times worse when using EO parameters based on a real-time trajectory, which is not acceptable for applications with the highest accuracy demands (i.e., ≤ 0.1 m in object space). It was also observed that the mapping accuracy is two to four times worse when based on direct orientation as compared to integrated sensor orientation.

Calibration

As soon as sensors such as GNSS are 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 directly measured coordinates are likely to occur. These differences cannot always be attributed to errors in the trajectory computations, especially when GNSS trajectory solutions can be delivered with high accuracy. Thus, such differences might also be caused by changes in the cam-

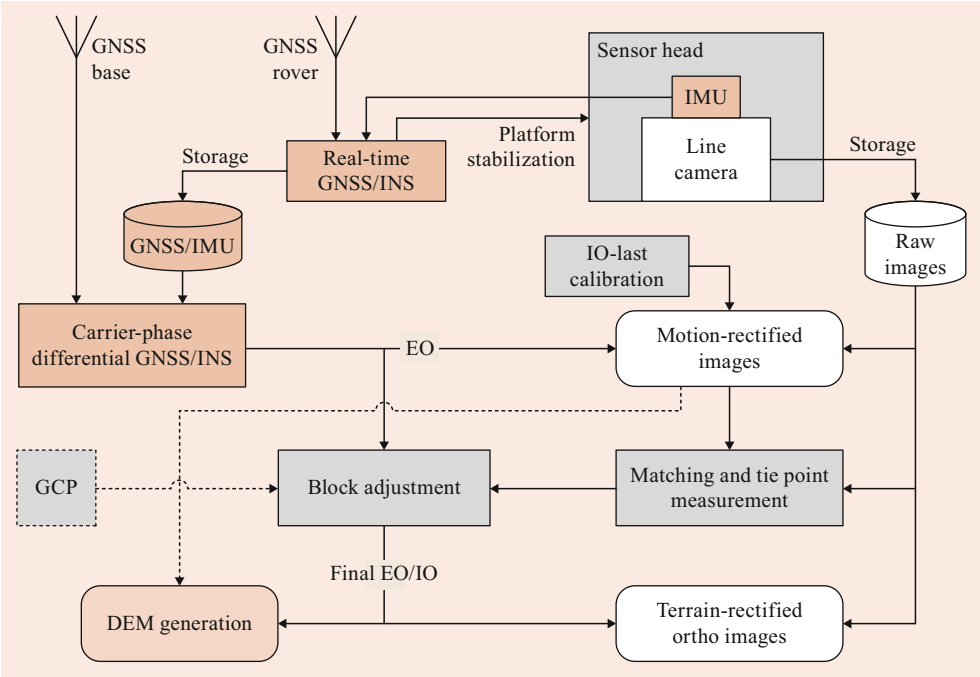

Fig. 9.37 Processing chain for line-camera data

era 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 Earth's 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 that directly relate to physical changes in the sensor geometry are well established in photogrammetric imaging [64] 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 [71] 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 the GNSS antenna and the inertial and camera perspective-center.

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 [42], the mitigation of residual systematic errors in trajectory determination [72], or both [39, 41, 73]. Such an adjustment process also serves 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 (based on return-intensity values). Contrary to cameras, this principle is not very suitable, as the corre-

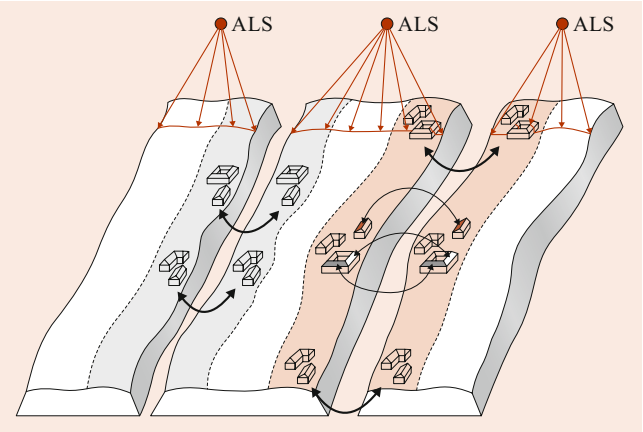

Fig. 9.38 Principle of surface-patch conditioning in airborne laser scanning. After [39]

spondence between laser points is only approximate. Modern approaches, therefore, rely on conditioning surface patches or other geometrical primitives that overlap between different passes (Fig. 9.38). 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 manmade structures. The parameters of the planes are estimated together with the calibration parameters that may include also biases in the trajectory [73]. Such trajectory bias modeling is approximate and can be avoided when the estimation includes the inertial raw readings as observations [68] (Sect. 9.4.6, *Angular Velocity and Specific Forces*). 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 (9.15), the observation equation for a laser target in the e-frame $\boldsymbol{x}_i^e$ lying on a plane $\boldsymbol{s}_j$ is given by

$$\left\langle \boldsymbol{s}_j \ , \begin{pmatrix} \boldsymbol{x}_i^e \\ 1 \end{pmatrix} \right\rangle = 0 \ , \tag{9.88}$$

where, the plane parameters are given by

$$\boldsymbol{s}_j = (s_1 \ \ s_2 \ \ s_3 \ \ s_4)^\mathsf{T} \ , \tag{9.89}$$

with s_1, s_2, s_3 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 $|\boldsymbol{s}_j| = 1$. The details of how such a constraint is added to the adjustment model are described in [42]. The principle can be extended to natural surfaces [74]; however, this approach has certain limits. First, most naturally flat surfaces are horizontal, which makes their contribution less significant. Second, perfectly flat surfaces are less common in nature, and their identification remains problematic [58]. 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 is the case of robotics, SLAM [59]. However, as terrestrial robots usually employ multibeam 3-D scanners, there is only one set of pose parameters per one instance of beam-array activation. This configuration is somewhat similar to that of line cameras and geometrically stronger than the more usual case of airborne scanning, where each single laser pulse has a unique set of poses.

9.5 Mapping Products

9.5.1 Surfaces

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.

Representation

The term digital elevation model (DEM) encompasses surface representation without specifying its nature. On the other hand, the 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 information captured about terrain height by means of optical sensors (Sect. 9.1) 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 of 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. Among the number of possibilities of terrain representation, grids, triangles, and contours are the most common and will, therefore, be discussed in more detail (Fig. 9.39).

Elevation grid

An elevation grid is the most straightforward representation of terrain (Fig. 9.39c). It is characterized by the regular, lattice organization of equally spaced points in the horizontal (x, y) projection. Each point of such a mesh contains one

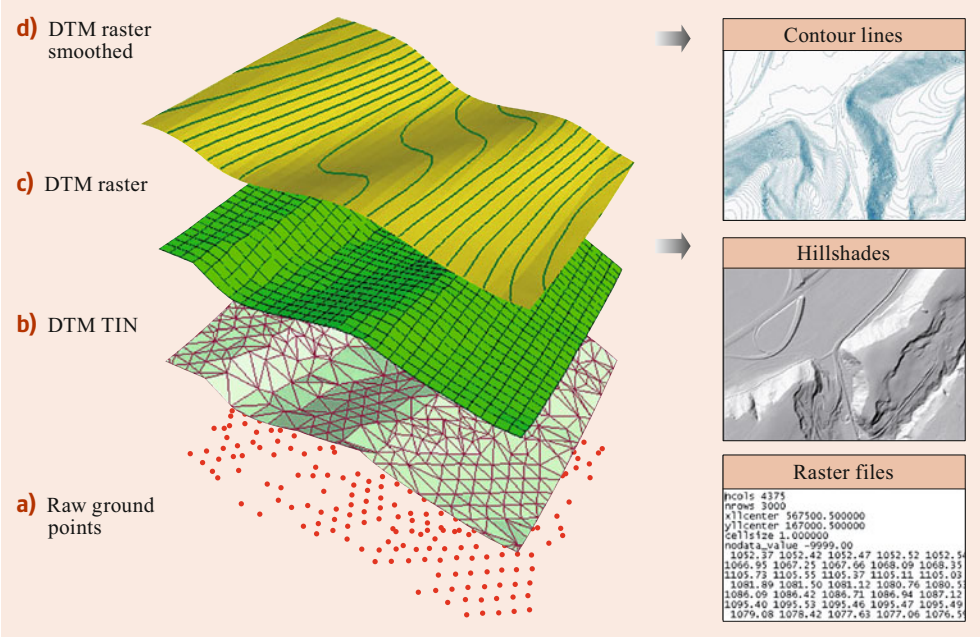

Fig. 9.39 Typical surface modeling based on irregularly distributed points (**a**). (**b**) Generation of a triangular irregular network (TIN). (**c**) Interpolation of TIN to regularly spaced grid. (**d**) Smoothing of grid and derivation of contour lines. After [12]

height (z) value for its location, which together with the x, y-coordinates is referenced to a common origin. The spacing between points is predefined and thus implies the resolution of the model. Such an organization is similar to that of an image and due to such a resemblance, this representation is referred to as a 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 of organization, which is also practical for further data manipulation. On the other hand, it is less suitable for the modeling of steep landscapes where the resolution with respect to slope normally decreases progressively (Table 9.13). It is also not suitable for the modeling of 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 suboptimal 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 an inconvenience could be circumvented by allowing the cell size to be adaptive (e.g., quadtree storage) or by applying image-like compression [75].

TIN

Compared to a grid, the triangulated irregular network (TIN) structure (Fig. 9.39b) is considerably more 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 fit in a mosaic that yields a surface. Such modeling is appropriate for areas with sudden changes in slope, where the edges of triangles can be aligned with discontinuities in the landscape (e.g., ridges, bottoms of gullies).

The storage requirements of irregularly spaced points and their 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 above-mentioned 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. 9.39d) was the most common way of surface modeling 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 onto a 2-D surface. In the past, contours were generated manually from oriented photographs on stereoplotters. 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 [75].

A comparison between different forms of terrain representation is given in Table 9.13.

Table 9.13 Comparison between different forms of DEM representation; ⊕ and ⊖ denote advantages and disadvantages, respectively

	Grid	TIN	Contours
Structure	⊕ Simple	⊖ Complex	⊖ Complex
Storage	⊕ More compact	⊖ Larger	⊖ Large
Exchange	⊕ Excellent	⊖ 2.5-D possible	⊖ Difficult
Applicability	⊖ Limited 2.5-D	⊕ 3-D possible	⊖ Limited 2.5-D
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

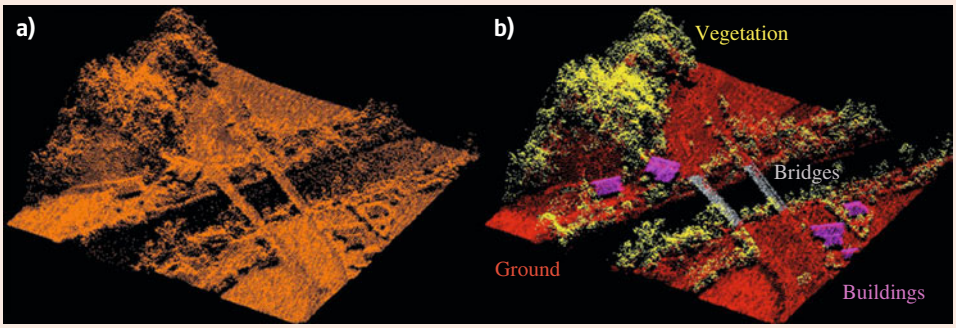

Fig. 9.40 Point-cloud classification, after [12]: (**a**) Raw laser point cloud, (**b**) point cloud classified into ground, bridge, vegetation, and building points

Reconstruction

DEMs of coarse resolution covering all continents are mostly produced by satellite missions (InSAR). The acquisition of DEMs 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 Sect. 9.4. Altimetric models of the highest precision and resolution usually come from airborne laser scanning.

Classification

As laser scanning is a nonselective 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. 9.40. This process is called classification and is highly but not entirely automated. When the point data are obtained by insertion from oriented images (Sect. 9.4), the classification can be performed together with image matching.

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. 9.39b). Triangulation can be performed in 2-D or in 3-D, according to the geometry of input data. Large, country-like elevation models are usually created 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. 9.41a) or complex manmade structures. As can be seen in Fig. 9.41b, performing 3-D triangulation is more suitable for this purpose, however, its evaluation is very complex in large datasets. Also, data exchange in the GIS community is not necessarily standardized for 3-D TIN structures.

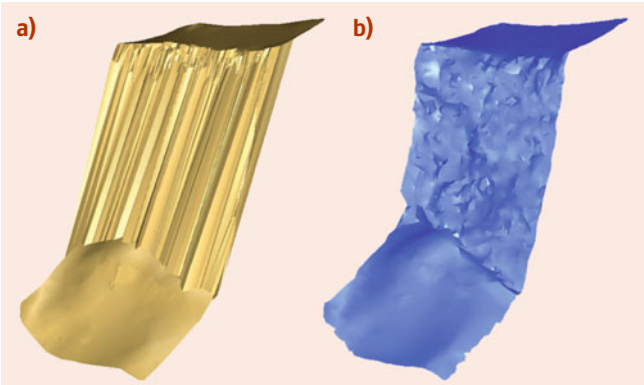

Fig. 9.41 Isometric view of a vertical rock face obtained by airborne laser scanning and triangulation; after [12]: (**a**) 2.5-D triangulation, (**b**) 3-D triangulation

Grid generation

An evenly-spaced DTM grid can be obtained by interpolation from the TIN facets. This approach is, however, only practical when the TIN model already exists. The grid can also be derived directly from the point data by various interpolation techniques. The popular interpolation methods used for this purpose are trend surface analysis, Fourier analysis, or kriging. These approaches have a global character

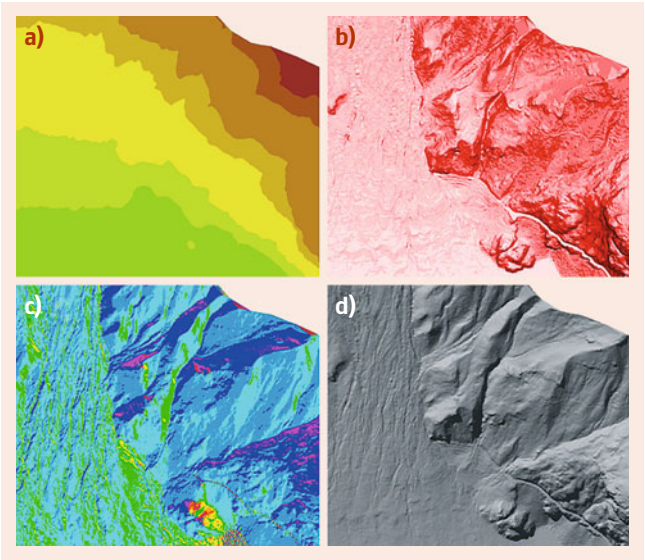

Fig. 9.42 Analysis of the raster height elevation model; after [12]: (**a**) DEM color coded by elevation, (**b**) DEM slope grid, (**c**) DEM color-coded by aspect, (**d**) DEM hillshade grid

and various levels of smoothing that are 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 [75].

In places where ground point sampling is low, like in forested areas or in 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.

DEM Analysis

Grid representation of DEM allows the application of image-like operations that are useful for highlighting different aspects of the surface. The most common are filters that perform terrain smoothing and gradient operations to visualize steepness and orientation, as shown in Fig. 9.42a–c. Fig. 9.42d depicts a so-called hillshade raster that improves visualization of the terrain by a chosen source (in position and angle) of illumination.

9.5.2 Orthophotos

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 a vertical view of 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 a central perspective. Frame cameras have a central perspective in x and y (Fig. 9.43a). Line cameras operating on satellite platforms, and in some cases also on aerial platforms, produce images that have a central perspective along the line photographed at one moment but an orthogonal perspective along the flight track (Fig. 9.43b).

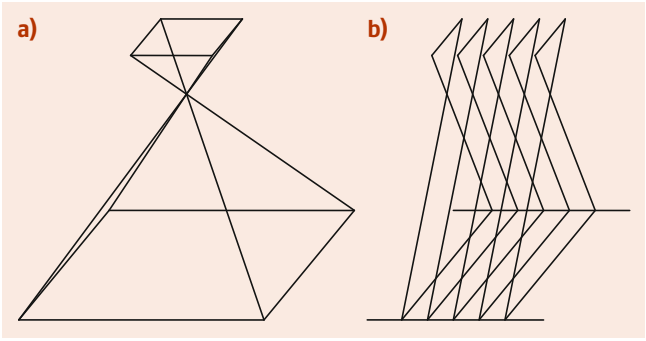

Fig. 9.43 Central perspective: (**a**) frame camera with central perspective in both dimensions, (**b**) line camera with central perspective across track and orthogonal perspective along track

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 this, the geometry of the camera plays a prominent role in the computation of an orthophoto.

Rectification Methods

The orthographic projection is obtained from the central perspective through an analytical process called rectification.

Table 9.14 Overview of the rectification methods

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)

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 Table 9.14.

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 a somewhat better two-dimensional relation between the image and the ground. Such an 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.

By far most orthophotos are produced by making use of DEM (Sect. 9.5.1), especially DTMs that exclude vegetation and manmade 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 DTM employed. A common procedure for orthophoto creation using DEM is the following:

1. An empty orthophoto is created at the start. Such an 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 of 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. 9.44). 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 necessarily correspond to the center of a pixel, its RGB-color (or gray) value on the orthophoto is found via interpolation with the neighboring pixels on the photograph. The interpolation can also be performed across several images, and these values can be further averaged to stabilize the resulting orthophoto radiometrically.

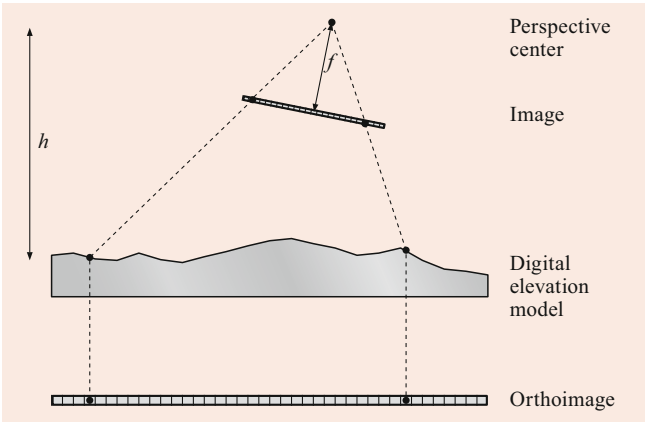

Fig. 9.44 The creation of an orthoimage with the use of DEM

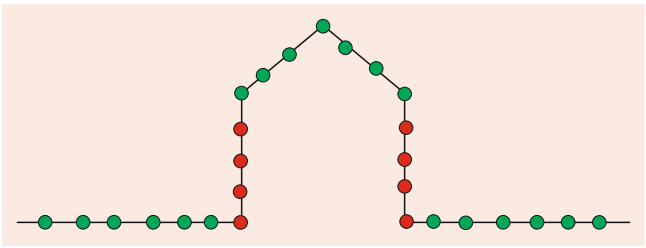

Fig. 9.45 Creation of a true orthophoto based on a dense point cloud (3-D). *Green points*: Points of the true orthophoto; *red points*: Points not used for the true orthophoto because they are not highest at their 2-D-position

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. 9.45).

References

1. Wolf, P.: Elements of Photogrammetry. McGraw-Hill, New York (1974)
2. Kraus, K.: Photogrammetry – Geometry from Images and Laser Scans. Walter de Gruyter, Berlin, Boston (2007)
3. Dorstel, C., Jacobsen, K., Stallmann, D.: DMC – Photogrammetric accuracy – Calibration aspects and generation of synthetic DMC images. In: Proc. Opt. 3-D Meas. Tech. Zurich, vol. VI, pp. 74–82. (2003)

4. Ladstadter, R., Tschemmernegg, H., Gruber, M.: Calibrating the ultracam aerial camera systems, an update. In: Proc. Int. Calibration Orientat. Workshop, EuroCOW, p. 8. (2010)
5. Gonzalez, R., Woods, R.: Digital Image Processing. Addison-Wesley, Reading (1992)
6. Ehlers, M., Klonus, S., Astrand, P., Rosso, P.: Multi-sensor image fusion for pansharpening in remote sensing. Int. J. Image Data Fusion **1**, 25–45 (2010)
7. Petri, G.: Deja vu – The configurations of the new airborne digital imagers are all rooted in the distant past! GeoInformatics **3**(5), 48–51 (2000)
8. Petri, G., Walker, S.: Airborne digital imaging technology: A new overview. Photogramm. Rec. **22**, 203–225 (2007)
9. Wehr, A., Lohr, U.: Airborne laser scanning an introduction and overview. ISPRS J. Photogramm. Remote Sens. **54**(2/3), 68–82 (1999)
10. Hug, C.: The scanning laser altitude and reflectance sensor – An instrument for efficient 3D terrain survey. In: The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 30, pp. 100–107. (1994)
11. Lohr, U., Beraldin, A., Blais, F.: Laser scanning technology. In: Airborne and Terrestrial Laser Scanning. Whittles Publishing, Dunbeath (2010)
12. Schaer, P.: In-flight Quality Assessment and Data Processing for Airborne Laser Scanning. PhD thesis. The Swiss Federal Institute of Technology, Lausanne (2009)
13. Carabajal, C., Harding, D., Luthcke, S.B., Fong, W., Rowton, S., Frawley, J.: Processing of shuttle laser altimeter range and return pulse energy data in support of SLA-02. In: The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 32-A4, pp. 269–277. (1999)
14. Spikes, V., Csatho, B., Whillans, I.: Airborne laser profiling of Antartic ice stream for change detection. In: The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 32-3-W14, p. 7. (1999)
15. Geiger, A., Kahle, H.G., Limpach, P.: Airborne Laser Profiling. ETH Research Database 6120. Swiss Federal Institute of Technology, ETH Zurich, Zurich (2009)
16. Morin, K.: Calibration of Airborne Laser Scanners. M.Sc.. Univ. Calgary, Calgary (2002). UCGE Report No. 20179
17. Glennie, C., Lichti, D.: Static calibration and analysis of the velodyne HDL-64E S2 for high accuracy mobile scanning. Remote Sens. **2**, 1610–1624 (2010)
18. Glennie, C., Kusari, A., Facchin, A.: 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. (2016) https://doi.org/10.5194/isprsarchives-XL-3-W4-55-2016
19. Betz, J.: Engineering Satellite-Based Navigation and Timing. Wiley, IEEE Press, Hoboken, New Jersey (2016)
20. Satirapod, C., Homniam, P.: GPS precise point positioning software for ground control point establishment in remote sensing applications. J. Surv. Eng. **132**(1), 11–14 (2006)
21. Titterton, D., Weston, J.: Strapdown Inertial Navigation Technology (1997). Part of IEE radar, sonar, navigation and avionics series, Stevenage, U.K., all the details in strapdown inertial navigation, attention: some mistakes in equastions
22. Jekeli, C.: Inertial Navigation Systems with Geodetic Applications. Walter de Gruyter, Berlin (2001)
23. Andersson, R., Bilger, H., Stedman, G.: Sagnac effect: A century of Earth-rotated interferometers. Am. J. Phys. **62**, 975–985 (1994)
24. Khaghani, M., Skaloud, J.: 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, vol. XLI-B3, pp. 827–831. (2016) https://doi.org/10.5194/isprsarchives-XLI-B3-827-2016
25. Khaghani, M., Skaloud, J.: Assesment of VDM-based autonomous navigation of a UAV under operational conditions. Robot. Auton. Syst. **106**, 152–154 (2018). minor review pending
26. Weiss, J., Kee, D.: A direct performance comparison between loosely coupled and tightly coupled GPS/INS integration techniques. In: Proc. 51st Annu. Meet. Inst. Navig., Colorado Springs, pp. 537–544. (1995)
27. Wei, M., Schwarz, K.P.: Testing a decentralized filter for GPS/INS integration. In: Proc. IEEE Symp. Position Locat. Navig. (1990)
28. Scherzinger, B.: Precise robust positioning with inertially aided RTK. J. Inst. Navig. **53**(2), 73–83 (2006)
29. Greenspan, R.: Inertial navigation technology from 1970 to 1995. J. Inst. Navig. **42**, 165–186 (1995)
30. Legat, K.: Approximate direct georeferencing in national coordinates. ISPRS J. Photogramm. Remote Sens. **60**, 239–255 (2006)
31. Skaloud, J., Legat, K.: Theory and reality of direct georeferencing in national coordinates. ISPRS J. Photogramm. Remote Sens. **63**, 272–282 (2008)
32. Rehak, M., Skaloud, J.: Fixed-wing micro aerial vehicle for accurate corridor mapping. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. II-1/W4, pp. 23–31. (2015) https://doi.org/10.5194/isprsannals-II-1-W1-23-2015
33. Cramer, M., Stallmann, D.: System calibration for direct georeferencing. In: ISPRS Comm. III Symp. Photogramm. Comput. Vis., Graz, p. 6. (2002)
34. Kruck, E.: Combined IMU and sensor calibration with BINGO-F. In: Proc. OEEPE Workshop Integr. Sens. Orientat. pp. 84–108. (2001). CD-ROM
35. Skaloud, J., Schaer, P.: Towards a more rigorous boresight calibration. In: ISPRS Int. Workshop Theory Technol. Real. Inert. GPS Sens. Orientat., Castelldefels (2003)
36. Cosandier, D.: Generating a Digital Elevation Model and Orthomosaic from Pushbroom Imagery. Univ. Calgary, Calgary (1999). UCGE Report No. 20133
37. Tempelmann, U., Hinsken, L.: Triangulation of ADS40 pushbroom image blocks – Not much different from classical frame blocks? In: SGPBF Annu. Meet. (2005)
38. Burman, H.: Calibration and Orientation of Airborne Image and Laser Scanner Data Using GPS and INS. PhD thesis. Royal Institute of Technology, Stockholm (2000)
39. Kager, H.: Discrepancies between overlapping laser scanning strips – Simultaneous fitting of aerial laser scanner strips. In: The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 35, pp. 555–560. (2004)
40. Morin, K., El-Sheimy, N.: Post-mission adjustment of airborne laser scanning data. In: FIG XXII Int. Congr. Washington, p. 12. (2002)
41. Friess, P.: Toward a rigorous methodology for airborne laser mapping. In: Proc. EuroCOW, p. 7. (2006). on CDROM
42. Skaloud, J., Lichti, D.: Rigorous approach to bore-sight self calibration in airborne laser scanning. ISPRS J. Photogramm. Remote Sens. **61**, 47–59 (2006)
43. Longuet-Higgins, H.C.: A computer algorithm for reconstructing a scene from two projections. Nature **293**, 133–135 (1981)
44. Hartley, R.I.: In defence of the 8-point algorithm. IEEE Trans. Pattern Anal. **19**(6), 580–593 (2012)
45. Schmid, C., Mohr, R., Bauckhage, C.: Evaluation of interest point detectors. Int. J. Comput. Vis. **37**(2), 151–172 (2000)
46. Jazayeri, I., Fraser, C.: Interest operators for feature-based matching in close range photogrammetry. Photogramm. Rec. **25**(129), 24–41 (2010)

47. Haralick, R.M., Shapiro, L.G.: Computer and Robot Visions. Addison-Wesley, Reading (1992)
48. Förstner, W., Gülch, E.: A fast operator for detection and precise location of distinct points, corners and centers of circular features. In: Proc. ISPRS Intercomm. Workshop Fast Process. Photogramm. Data, pp. 281–305. (1987)
49. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proc. 4th Alvey Vis. Conf. pp. 147–151. (1988)
50. Lowe, D.G.: Distinctive image features from scale invariant key points. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
51. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell. **27**(10), 1615–1630 (2005)
52. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications ACM **24**, 381–395 (1981)
53. Nistér, D.: An efficient solution to the five-point relative pose problem. IEEE Trans. Pattern Anal. Mach. Intell. **26**(6), 756–770 (2004)
54. Szeliski, R.: Computer Vision: Algorithms and Applications. Springer, London (2010)
55. Hirschmüller, H.: Stereo processing by semiglobal matching and mutual information. IEEE Trans. Pattern Anal. Mach. Intell. **30**(2), 328–341 (2008)
56. Haala, N.: Dense Image Matching Final Report, pp. 115–145 (2014). Tech. Rep., EuroSDR Official Publication No. 64
57. Colomina, I., Blazquez, M.: A unified approach to static and dynamic modeling in photogrammetry and remote sensing. In: The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 35-B1, pp. 178–183. (2004)
58. Kerstling, A.P., Habib, A., Bang, K.I., and Skaloud, J.: Automated approach for rigorous light detection and ranging system calibration without preprocessing and strict terrain coverage requirements. Opt. Eng. **51**(7), 076201 (2012)
59. Strasdat, H., Montiel, J., Davison, A.: Real-time monocular slam: Why filter? In: IEEE Int. Conf. Robot. Autom. (ICRA), pp. 2657–2664. (2010)
60. Cucci, D.A., Rehak, M., Skaloud, J.: Bundle adjustment with raw inertial observations in UAV applications. ISPRS J. Photogramm. Eng. Remote Sens. **130**, 1–12 (2017)
61. Freyer, J., Brown, D.: Lens distortion for close-range photogrammetry. Photogramm. Eng. Remote Sens. **52**, 51–58 (1986)
62. Gruen, A.: The accuracy potential of the modern bundle block adjustment in aerial photogrammetry. Photogramm. Eng. Remote Sens. **48**, 45–54 (1982)
63. Rehak, M., Skaloud, J.: Time synchronization of consumer cameras on micro aerial vehicles. ISPRS J. Photogramm. Remote Sens. **123**(1), 114–123 (2017)
64. Förstner, W., Wrobel, B.: Mathematical concepts in photogrammetry. In: McGlone, J., Mikhail, E. (eds.) Manual of Photogrammetry, 5th edn., pp. 15–180. ASPRS, Bethesda (2004)
65. Bjerhammar, A.: Theory of Errors and Generalized Matrix Inverses. Elsevier, Amsterdam (1973)
66. Kummerle, R., Grisetti, G., Strasdat, H., Konolige, K., Burgard, W.: g^2o: A general framework for graph optimization. In: IEEE Int. Conf. Robot. Autom. (ICRA), pp. 3607–3613. IEEE, New York (2011)
67. Davis, T.A.: Direct Methods for Sparse Linear Systems. SIAM, Philadelphia (2006)
68. Cucci, D., Skaloud, J.: On inertial measurements in dynamics networks. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. IV-2/W5, pp. 549–557. (2019)
69. Hoffman, O., Nave, P., Ebner, H.: DPS – A digital photogrammetric system for producing digital elevation models and orthophotos by means of linear array scanner imagery. In: The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 24, pp. 216–227. (1982)
70. Cramer, M.: The ADS40 Vaihingen/Enz geometric performance test. ISPRS J. Photogramm. Remote Sens. **60**, 363–374 (2006)
71. Ebner, H.: Self calibrating block adjustment. In: XIIIth Congr. Int. Soc. Photogramm. Com. V., Stockholm (1978)
72. Filin, S., Vosselman, G.: Adjustment of airborne laser altimetry strips. In: ISPRS XXth Congr. Geo-Imagery Bridg. Cont. pp. 285–289. (2004)
73. Glira, F., Pfeifer, N., Mandelburger, G.: Hybrid orientation of airborne lidar point clouds and aerial images. In: he International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. IV-2/W5, pp. 567–574. (2019)
74. Filin, S.: Recovery of systematic biases in laser altimetry data using natural surfaces. Photogramm. Eng. Remote Sens. **69**(11), 1235–1242 (2003)
75. El-Sheimy, N., Valeo, C., Habib, A.: Digital Terrain Modeling. Artech House, Norwood (2005)

Jan Skaloud Jan Skaloud is a senior scientist and lecturer at the Swiss Federal Institute of Technology Lausanne (EPFL) and Adjunct Professor in the Department of Geomatics Engineering at The University of Calgary. His research is related to satellite and inertial navigation, sensor calibration, and modern mapping techniques with active and passive optical sensors. He is a member and active participant in numerous international scientific organizations in this field.

Michael Cramer Michael Cramer is Senior Lecturer at the Institute for Photogrammetry (ifp), University of Stuttgart, Germany. He received his Doctor's degree in Geodesy and Geoinformatics from the same university. His research was previously devoted to georeferencing of sensors using integrated GPS/inertial data for airborne photogrammetry. Currently, his focus is on digital photogrammetric airborne sensor systems with special emphasis on their overall accuracy. He also is active in standardization activities worldwide.

Norbert Haala Norbert Haala is Associate Professor at the Institute for Photogrammetry, the University of Stuttgart, where he is responsible for lectures in the field of photogrammetric image processing. His research interests include virtual city models and image-based 3-D reconstruction.