Introduction to Photogrammetry

Pensum for GEO4530

UiO : Universitetet i Oslo

Luc Girod
Department of Geosciences
Faculty of Mathematics and Natural Sciences
University of Oslo

Oslo, Norway
April 2019

Contents

Photogrammetry : from the greek *photos* meaning *light*, *gramma* meaning *something drawn or written*, and *metron* meaning *to measure*, it is the science of making measurements from photographs [McGlone et al., 2004].

1 Introduction

1.1 Main principle

The idea behind photogrammetry is to use the difference in perspective in images taken from two different positions to compute 3-dimensional information, mimicking the human depth perception.

The fundamental concept is simple (shown in Fig. 1). Suppose you have two images (Im_1 and Im_2), and you have information on (1) the internal characteristics of the camera(s) used to take them, (2) the location from where the pictures were taken ($CameraPosition_1$ and $CameraPosition_2$), and (3) how the camera was oriented in space then. If it is possible to identify, manually or automatically, an object A in both images (points a_1 and a_2), you can obtain the 3D position of the object by computing the coordinates of the intersection of the projective rays from each image corresponding to the image of object A (($a_1 - CameraPosition_1$) and ($a_2 - CameraPosition_2$)). By applying this principle to all identifiable points, it is possible to reconstruct the three-dimensional shape of a photographed object or scene.

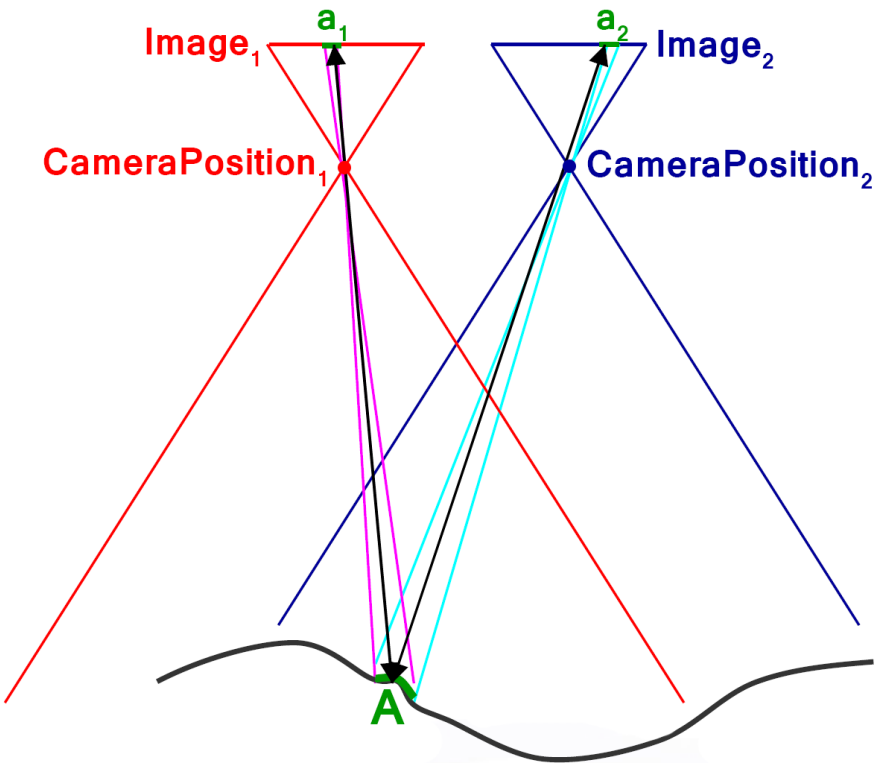

Figure 1: Fundamental concept of photogrammetry.

1.2 Elements of photography

The first imaging systems were developed in the mid 1800s (see Fig. 2, [Niépce, 1839]) and consisted of a box with a pinhole and a surface coated with light-sensitive chemicals (often silver based). The invention of lens cameras and the use of more sensitive chemicals for films – and later digital sensors – improved dramatically the amount of light the systems could capture in a relatively short time, hence making the process of photography a very practical way of "capturing a slice of reality".

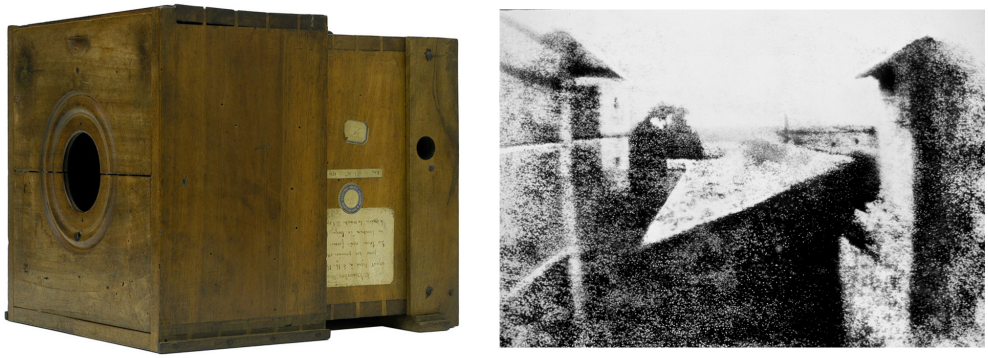

Figure 2: Left: The first working camera used by Nicéphore Niépce. Right: "Point de vue du Gras" - The first picture ever taken - Nicéphore Niépce - 1826-7.

1.2.1 Architecture of a modern camera

The fundamental structure of cameras has not changed much in the last decades, except for the replacement of the photographic film by a digital sensor . The key elements composing a modern camera are shown in Figure 3. The characteristics of each element affects the camera parameters (see Section 1.2.2) hence the image.

1.2.2 Camera lens and sensor parameters

Modern cameras have increasingly complex optical systems, using various types of glass and glass coating, combining a number of lenses of different type (convergent, divergent, aspherical...) to obtain increasingly polyvalent optics, with lower levels of geometric and chromatic distortions. The following sections will explore the effect of the camera parameters on the images taken.

1.2.2.1 Perfect camera optical parameters

A camera has four critical parameters:

- The focal length **Foc** (in mm). It influences the zoom level: a longer focal (Foc is a larger number) will result in a narrower field of view (see Figure 4).
- The aperture (or f number). It influences both the amount of light going through the lens and the depth of field (see Figures 5 and 6). It also influences the vignetting (fall-off of the brightness away from the image center, see Fig. 7).
- The exposure time. It is the time duration during which the light can reach the film/sensor.
- The sensitivity, also typically called **ISO**. It is a measure of how reactive the film/sensor is to light stimulation, a higher ISO number indicates that less light is needed to obtain the same brightness in the image, but also comes with increased noise.

In Figure 6, the three images have the same level of luminosity, since the exposure time was changed along the aperture. The brightness of an image follows the following formula:

$$Brightness \propto \frac{ExposureTime * ISO}{Aperture^2}$$

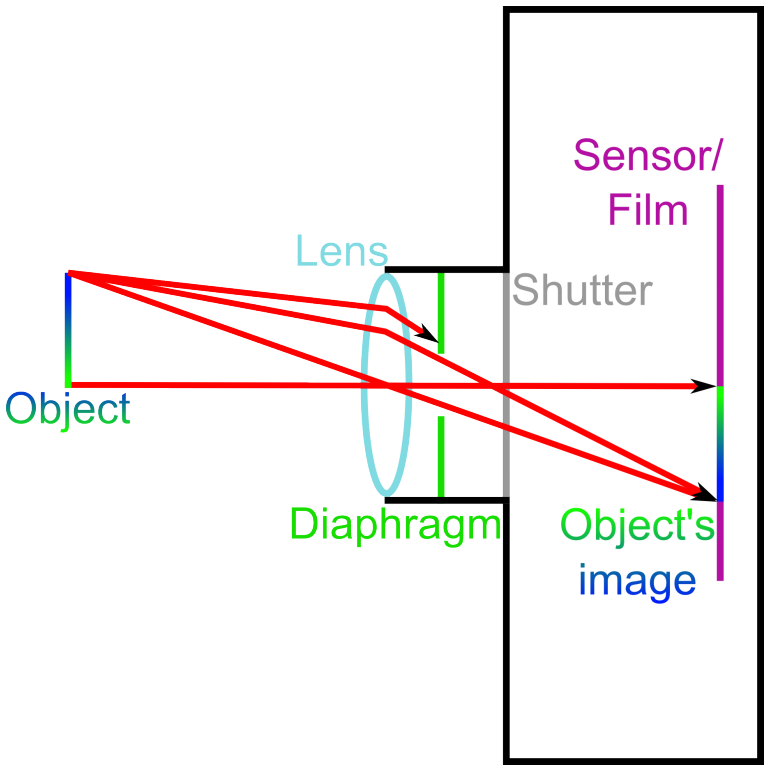

Figure 3: Simplified diagram of a camera with an object in the focus plane.

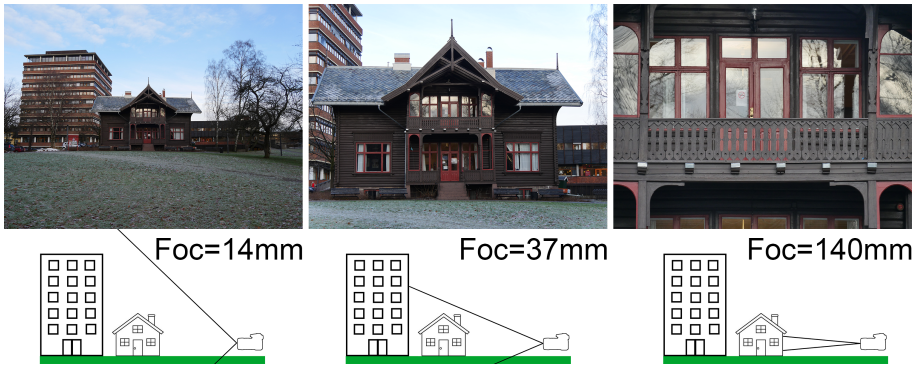

Figure 4: Images taken from the same point with increasing focal lengths, showing more, respectively less, of the scene with lower, respectively higher, levels of detail.

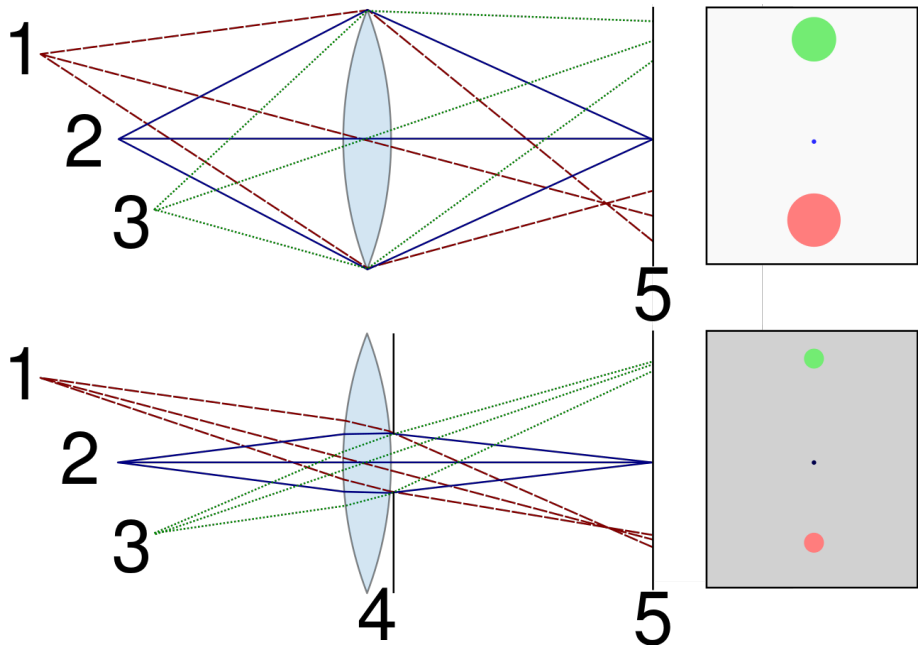

Figure 5: Diagram of the effect of aperture on Depth of Field – top: wide open (small f number) ; bottom: closed aperture (high f number) – 1: point farther than focus plane ; 2: point in focus plane ; 3: point closer than focus plane ; 4: diaphragm ; 5: sensor/film.

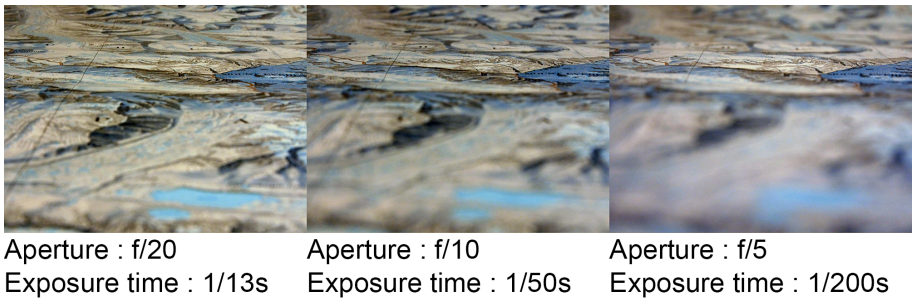

Figure 6: Example of the effect of aperture on Depth of Field. Note that the exposure times compensate for the different amount of light let through by the aperture blades to provide similar brightness.

Figure 7: Vignetting effect on an homogeneously gray, evenly lit surface (Nikon D90 with Nikkor 18-105 VR at 18mm-f/3.5)

1.2.2.2 Distortion

Distortions are divergence from the perfection of a theoretical camera. There are two types of distortion:

- Geometric distortion result of the imperfection of the optical system and flatness of the sensor. It is more visible in zoom lenses (variable focal) and is extreme in fish-eyes. The most noticeable effect is the abnormal curvature of straight lines (see Figure 8).
- Chromatic distortion, resulting of the variation of refractive index depending on the wave-length (higher for blue than red, see Figure 9).

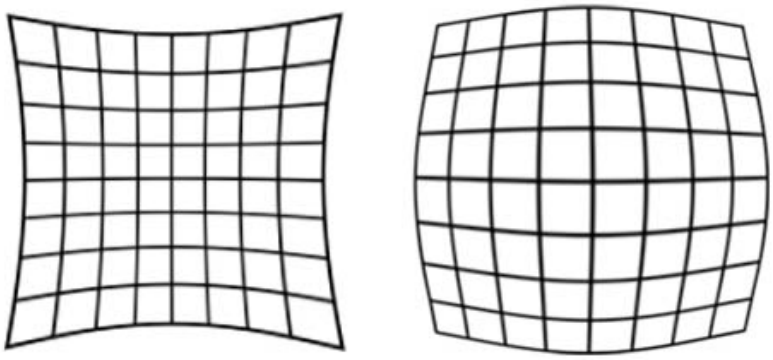

Figure 8: Radial geometric distortion of a regular grid. Left: pincushion – Right: barrel.

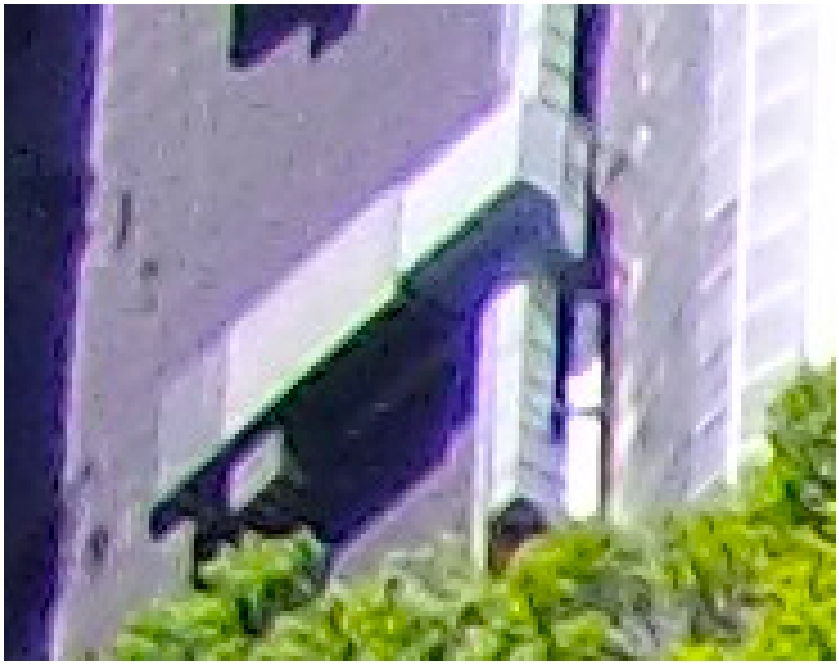

Figure 9: Example of chromatic distortion with green and purple fringes in sharply contrasted areas.

1.2.3 Digital camera sensors

Historically, the light was captured on film (mostly silver based solutions), but the first digital sensors was developed in 1975 by [Lloyd and Sasson, 1978] at Eastman Kodak, and had a resolution of 100*100 pixels.

A sensor is a matrix of light sensitive cells called photo sensors, that convert the light that strikes them into electrical current. The current stored in each cell is then converted into digital information. Different technologies with different advantages and disadvantages exist. CCD used to be prominent but is only found in very big sensors these days, being replaced by CMOS type sensor and variants. CMOS offers faster readout time, lower power consumption and are less expensive to produce.

An array of photo sensitive sensors only creates a black and white picture, so additional systems are required to get coloured images. The most common is the Bayer matrix (see Fig. 10, left): a layer of colour filters is laid on top of the sensor so each photo sensitive cell only receives a certain range of light frequency. There is twice as many green tiles as blue or red because the typical human eye (not affected by colour-blindness or quadri-chromatism) is more sensitive to green (see Fig. 10, right) and therefore the higher accuracy in the green wavelength is beneficial to the perceived image quality. Then the colour for each pixel is interpolated from the neighbouring pixels.

Other patterns exist (like on the Fujifilm X-trans sensors) as well as other type of sensors like the Foveon X3 that is using several layers of sensors to capture full resolution images in each of the wavelength bands. Some systems are even using colour separating prisms to send different wavelength to separate sensors, or different lenses for each colour, and fuse the resulting images together in software.

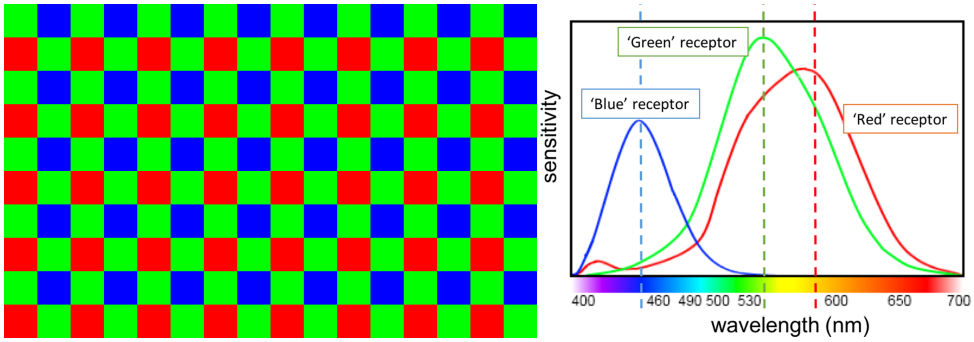

Figure 10: Left: Typical Bayer matrix; Right: sensitivity of the different receptors ("Blue", "Green" and "Red") of the typical human eyes (from https://people.eecs.berkeley.edu/~cecilia77/graphics/a6/).

1.2.4 Sensors capturing a wider range of the electromagnetic spectrum

Cameras were first developed to capture images of the world as humans see it. However, the electromagnetic spectrum is not limited by human biology, and the observation of other wave lengths can provide additional information. For instance, the near infra-red wavelength (NIR) is very useful to identify vegetation because plants reflect it strongly. A great way to detect vegetation is through the normalized difference vegetation index (NDVI), a combination of the red and near infra-red bands. Fig. 11 shows the section of the electromagnetic spectrum used in optical remote sensing.

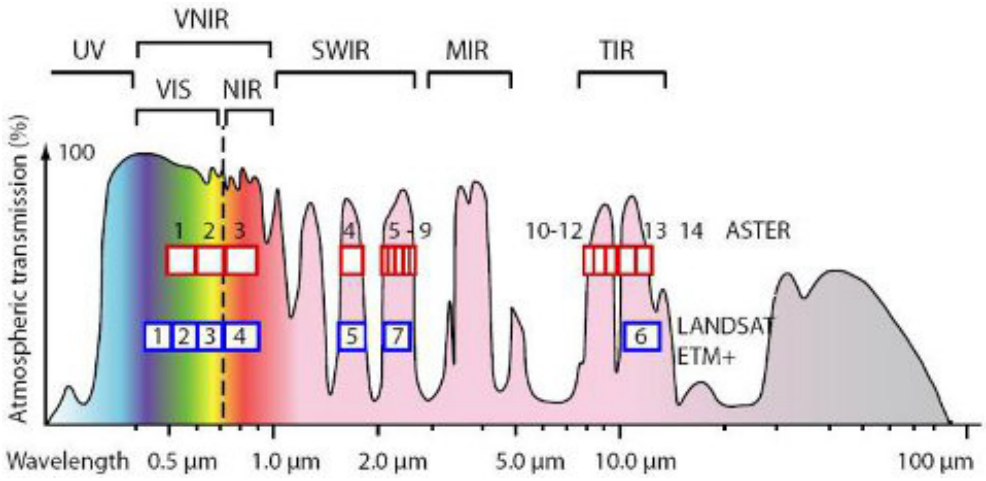

Figure 11: The part of the electromagnetic spectrum used in optical remote sensing with indications of the typical subdivisions, as well as indications of the discrete limits of the bands from the ASTER and LandSat ETM+ instruments (see Section 1.2.6). In the background is the standard atmospheric transmission of signals in each wavelength. Figure from [Kääb, 2005].

1.2.5 Video cameras

Soon after the invention of photography came cinema. A video camera is simply a camera that can take pictures in fast succession (the typical rate for cinema is 24 frames per second, for TV 25 or 29.97 depending on the standard, but other, higher values are available today).

1.2.6 Pushbroom cameras

Pushbroom cameras, also called digital scanners are cameras that replace the bi-dimensional sensor of typical cameras with a mono-dimensional, linear one [Gupta and Hartley, 1997]. The image is then acquired by moving the camera in the direction perpendicular to the linear array of sensors. This affects the geometry of the image in a number of ways, the most notable one is the perspective and distortion that are not radial anymore, but in a repeating pattern for each line of data (see Fig. 12).

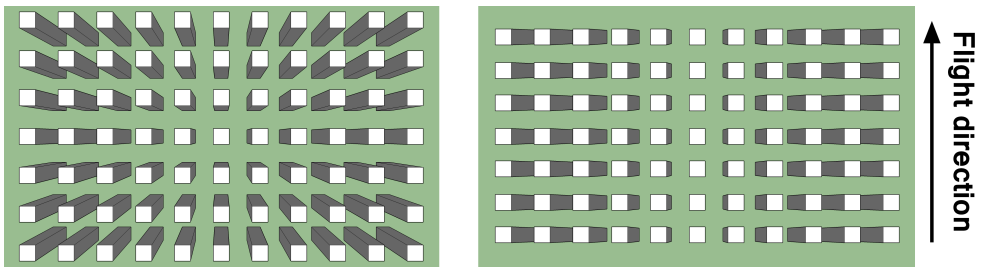

Figure 12: Foreshortening for a frame camera (left) and for a pushbroom camera (right).

1.2.7 Stereo cameras

If most camera systems take a single image at a time, some can take multiple. This idea is usually implemented in camera systems composed of several cameras rigidly linked together (possibly in the same camera body), pointing in the same direction, hence creating an instantaneous (multi-)stereo image set. Another kind of system can also be called stereo by some: stereo camera taking images simultaneously, but pointing in different directions. Here the idea is that the camera system is in movement and that the area imaged by one of the cameras a time t_1 is going to be imaged by an other camera at time t_2: this is the kind of camera used for along-track stereo imaging.

1.3 A short history of Photogrammetry

Photogrammetry is a fairly modern science that developed in parallel with photography. If the word photogrammetry itself was used in print for the first time in [Meydenbauer, 1867], the science itself started nearly half a century earlier with the works of Aimé Laussedat, a colonel in the French Army Corps of Engineers [Laussedat, 1854]. The first use of photography for the acquisition of topography was through terrestrial photography, images taken from the ground, as a mean to enhance and increase the density of the data acquired with theodolites. Photogrammetry did however quickly take to the skies with the use of kites and hot air balloons at first, and then with the invention of airplanes.

As the photography related technologies – optics and film chemistry – evolved, photogrammetry evolved as well, with better tools leading to easier and more accurate methods to extract the data from the imagery. Specific instruments for photogrammetry were developed, such as Poro's photogoniometer (1865), Deville's stereo-planigraph (1896) or the first purposed designed planned mounted camera by the Brock brothers (1914). In 1921, Fairchild created a mosaic of 100 images over the island of Manhattan, before inventing the gyro-stabilized camera [Fairchild and Morton, 1928]. More efficient tools to derive cartography (more specifically contour lines for topographic maps) were developed in the form of stereoplotters (see Fig. 13) starting in the 1930s, gradually gaining complexity and precision, as well as improving the ease of use. The use of photogrammetry for military reconnaissance and general topography only grew in popularity going forwards [Saint-Amour, 2011].

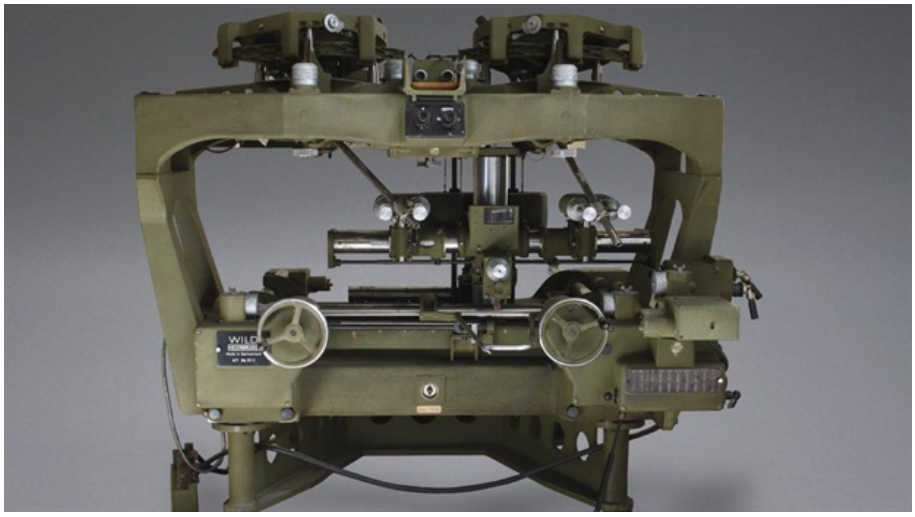

Figure 13: Autograph Wild A7 photogrammetric stereoplotter (Image courtesy of the Technical Museum Vienna).

1.4 Computer enabled photogrammetry

The computer revolution had a large impact on every aspect of photogrammetry, from acquisition, processing, storage and even possible data product with the development of gridded digital elevation models (DEMs, see Section 2.2) and orthorectification (see Section 2.4).

The first step consisted in bringing the data captured by stereoplotters into the computer for an easier storage and data manipulation. Then, the images themselves were brought into the computer, first through scanning and then through digital photography,

allowing for faster and more streamlined visualization and interaction (changing the visualized couple of images would take a simple click instead of a complicated manipulation off the stereoplotter). Relative and absolute orientation data (position and viewing angles of the camera) could then also be solved analytically by imputing the positions of tie points (TPs) and ground control points (GCPs) and solving the systems of equations.

The development and availability of highly precise and accurate ground Global Navigation Satellite Systems (GNSS) systems, such as the GPS, first allowed for an easier process for gathering GCPs, and then, with embarked GNSS systems, for a gradual decrease in the amount of GCPs required for georeferencing.

The automation of a number of processes came quickly afterwards. In modern photogrammetry, the detection of tie points (through algorithms such as SIFT [Lowe, 2004]) as well as the computation of elevation data (through dense correlation) is fully automated. Professional photogrammetric surveys are now typically processed with no human input after the acquisition of the image and orientation data.

In parallel, photogrammetry became more accessible to non-experts and compatible with non-specialized, relatively cheap hardware. A lot of this advancement can be attributed to the process of Structure-from-Motion (SfM) presented by [Koenderink and Van Doorn, 1991] and first implemented into efficient, publicly available code by [Snavely et al., 2006; Snavely, 2010] in the Bundler package. SfM allows for the automatic computation of both the relative external orientations (positions and view angles in a relative space) and internal orientations (also called camera calibration, the information about the cameras' optics and sensors) of a group of images without *a priori* knowledge when tie points can be identified (also automatically). The information provided by SfM can then be used for multiview stereo to robustly and automatically compute accurate and dense 3D models [Furukawa and Ponce, 2010]. Better and faster implementations have since been developed, both in the form of commercial software (for instance Agisoft Photoscan [Agisoft LLC, 2017] or Pix4D [Pix4D SA, 2017]) and in the form of open-source projects such as MicMac [Pierrot-Deseilligny et al., 2017; Rupnik et al., 2017], the software used throughout this coursework.

2 Output products

Photogrammetry can produce a number of different products for different applications, a few examples are described in this section.

2.1 Topographic maps

Creating topographic maps was the first aim of photogrammetry. Using stereo-data, contour lines could be drawn, creating invaluable cartographic information for a number of applications, most notably military.

2.2 Digital Elevation Models

A Digital Elevation Model (DEM) is a georeferenced grid associating height values to each position of the grid, where the distance between two cells of the grid is called **resolution** or **Ground Sampling Distance (GSD)**, see Fig.14. It is the logical evolution of topographic maps in the digital age, presenting a much higher density of directly accessible information. A very common file format to store DEMs is a variation on the raster graphics format tiff, the GeoTiff. This format is simply adding geographic metadata to a tiff file.

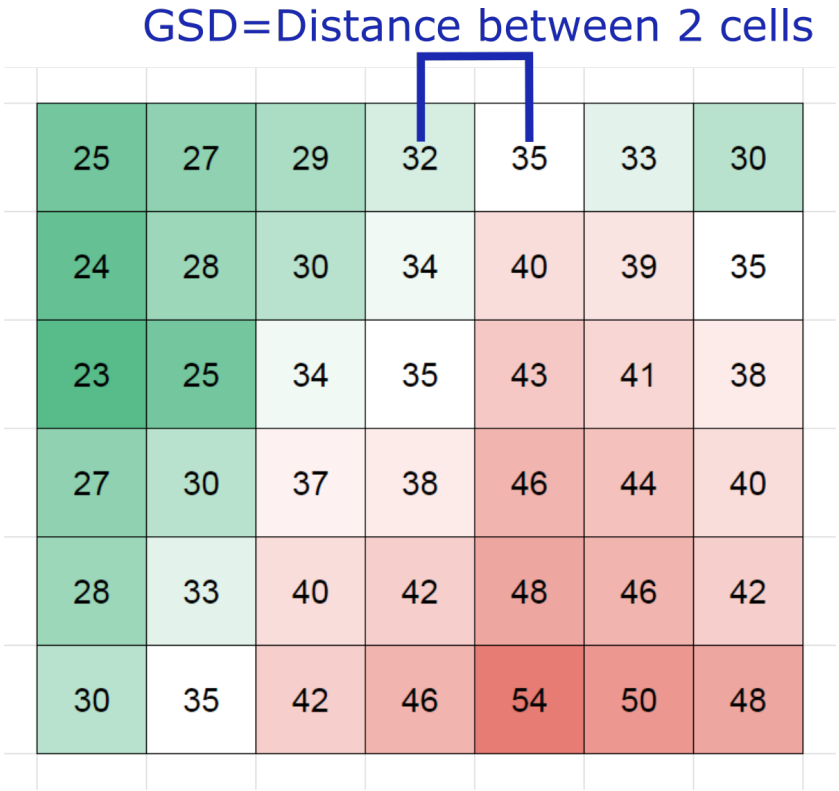

Figure 14: DEM and GSD.

Several sub-categories of DEM exist (see Figure 15 for a graphical view):

- A DSM (Digital Surface Model) represents the terrain variability and the objects (or superstructures) on top of it.
- A DTM (Digital Terrain Model) is a sub-product of a DSM where all the building, trees and small objects are removed to only represent the terrain .

- A full 3D description of the terrain also describes information about the potential overhangs such as the terrain bellow a bridge or an overhanging cliff, as well as the facades of buildings. Such data cannot be stored in a single layer grid.

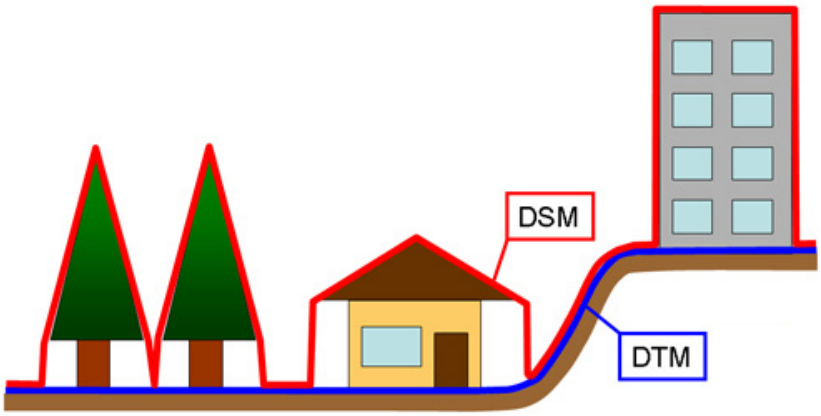

Figure 15: Illustration of the difference between DSM and DTM.

2.3 Differential Digital Elevation Models

A Differential Digital Elevation Model dDEM is the comparison of two DEMs showing the difference in elevation between the two. They can be used to assess the quality of one of the DEMs when compared to validation data, to be a first estimation of a georeferencing bias, or to investigate change in the topography (see Fig. 16) caused by e.g. glacier thinning or landslides.

2.4 Orthoimages ("pseudo" or "true")

An orthoimage is an image geometrically corrected (through a process called orthorectification) for scale variations induced by topography (see Figure 17). It is usually projected into a map reference system and therefore overlay-able to a map. Like a DEM, it is a grid with a defined GSD, see Fig.14.

We call an orthoimage "true" when it is made using a full resolution DSM, allowing for the geometric correction of buildings and other superstructures but running the risk of having data voids if some part of the ground was not imaged at least twice because of occluding objects. A "pseudo" orthoimage is computed using a DTM to correct the images, creating images with visible foreshortening and hidden parts by the ignored superstructures.

The process of orthorectification is explored in Section 3.10.

2.5 Orthophotomosaic

An orthophotomosaic is a mosaic of orthoimages covering a larger area than a single orthoimage could. Except in the case of satellite imagery where single images cover very large areas, most products describbed as orthophotos are actually orthophotomosaics, as a single orthorectified image usually covered only a small area. Section 3.11 explains the process of mosaicking.

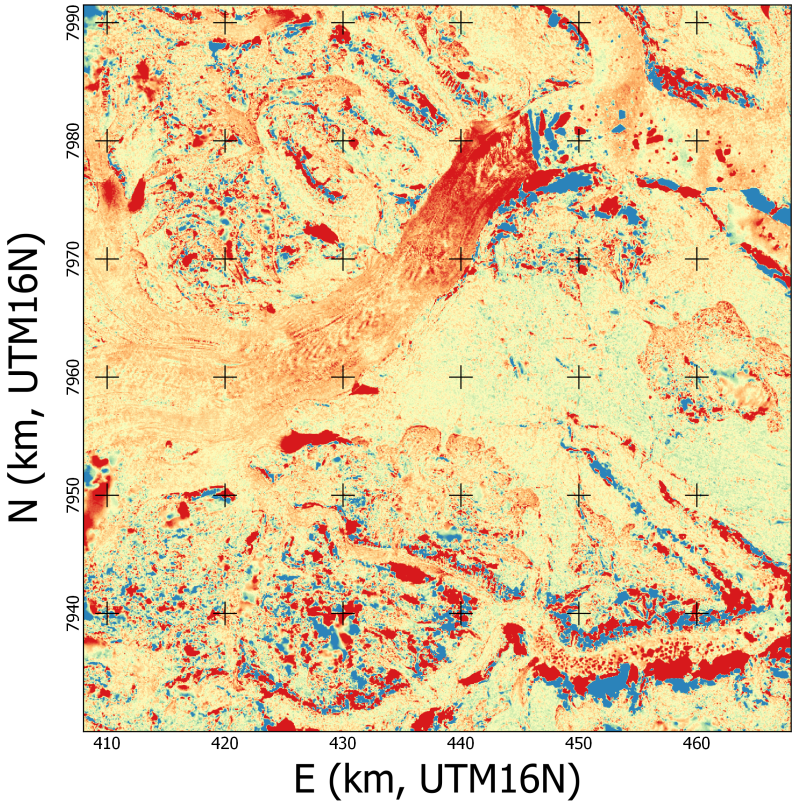

Figure 16: dDEM between two SPOT-5 HRS DEM from 2007 and 2014 (acquired for the SPIRIT program [Korona et al., 2009]) over the Daugaard-Jensen outlet glacier in eastern Greenland showing elevation change due to the thinning of the ice.

2.6 Thematic maps

Since an orthoimage is overlay-able to a map, it is possible to create a thematic map out of an orthoimage and add topographic data from a DEM. The orthoimage is used to identify roads, buildings, fields, forests, rivers and other features of interest that will then be added to the map. This process can be done manually or automatically by using remote-sensing and computer visions methods to classify the image.

2.7 3D models

By using the convergent method (see Sections 3.1.1 and 3.9.2) or the combination of nadir and oblique aerial photography, it is possible to compute a textured full 3D model of a scene. Fig. 18 shows such a 3D model as well as some of the pictures used in the computation.

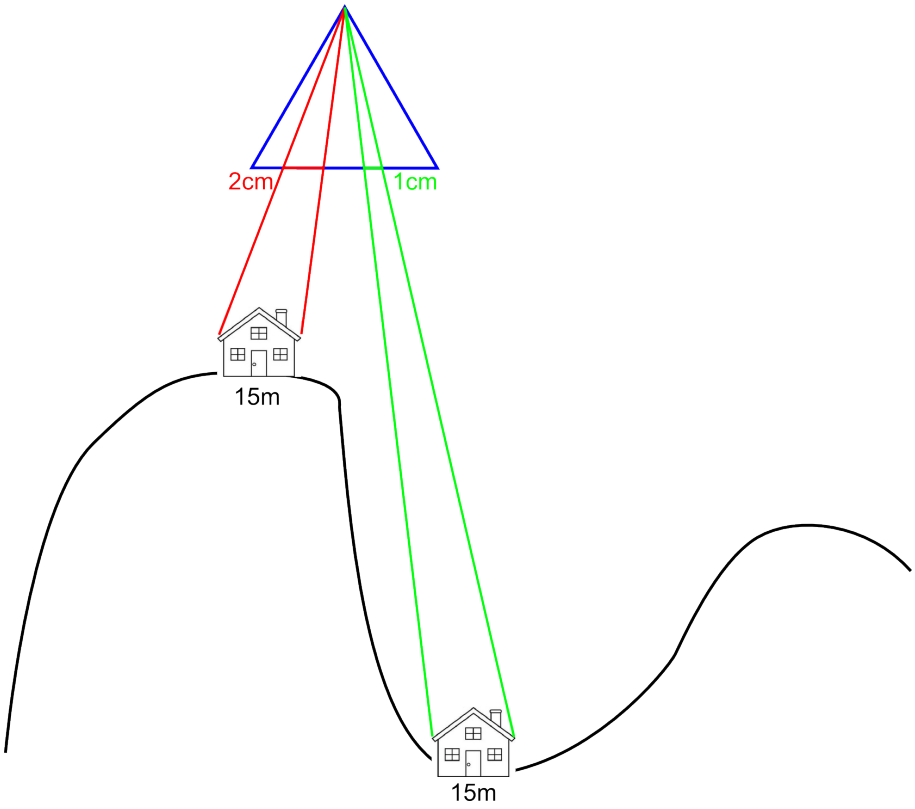

Figure 17: Illustration of the scaling problem in non-rectified images.

Figure 18: Sample images for the computation of a 3D model of a miniature house and resulting model - from [Girod and Pierrot-Deseilligny, 2014].

3 The photogrammetric processing chain

The modern, computerized, structure-from-motion (SfM) enabled photogrammetric process can be divided in different steps, as shown in the figure 19.

3.1 Image acquisition

There are two main methods of acquisition in photogrammetry, the *convergent method* (see Section 3.1.1) and the *parallel method* (see Section 3.1.2).

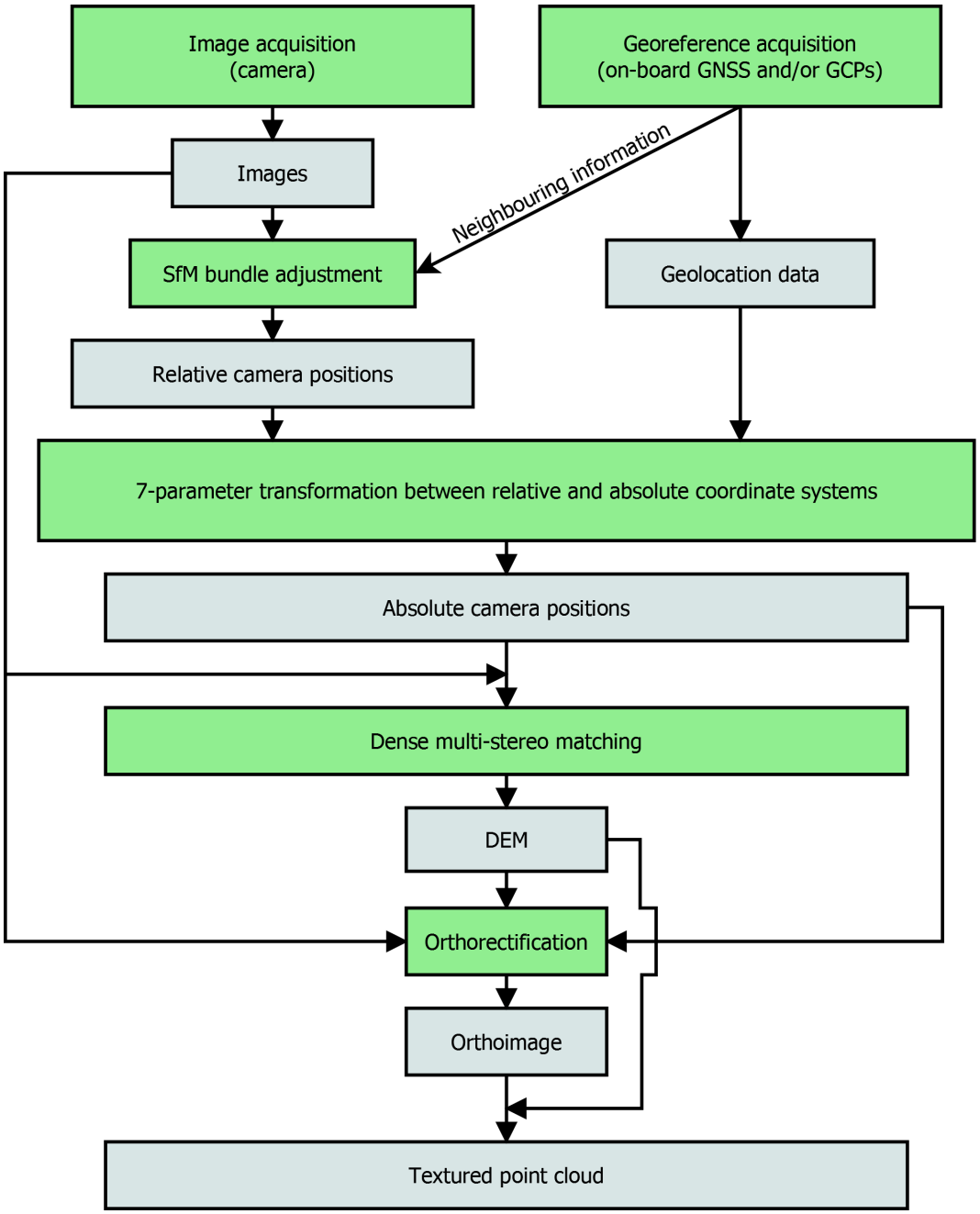

Figure 19: The SfM photogrammetric workflow - Input data and products are in grey and processes in green.

3.1.1 The convergent method

In this first method, pictures are taken aiming at the same point in space. The primary picture in the middle will take advantage of the secondary ones to get multi-stereoscopy, and therefore 3D information. A simple setup is seen in Fig. 20A.

If a single point of view is not sufficient to see the whole object, it is possible to take other sets of images to cover the scene. Fig. 20B shows an example of setup for a single

plane 360 degrees view of an object, and a setup with several circles taken from different altitudes is the logical next step. Having linking images ensures a robust geometrical link between the different points of view.

The angle between two lines of view must be at the same time sufficient to provide stereoscopy (very small angles create higher incertitude) and small enough for both the computation of tie points and correlation to work. A rule of thumb is that a good value is between 10 and 15 degrees.

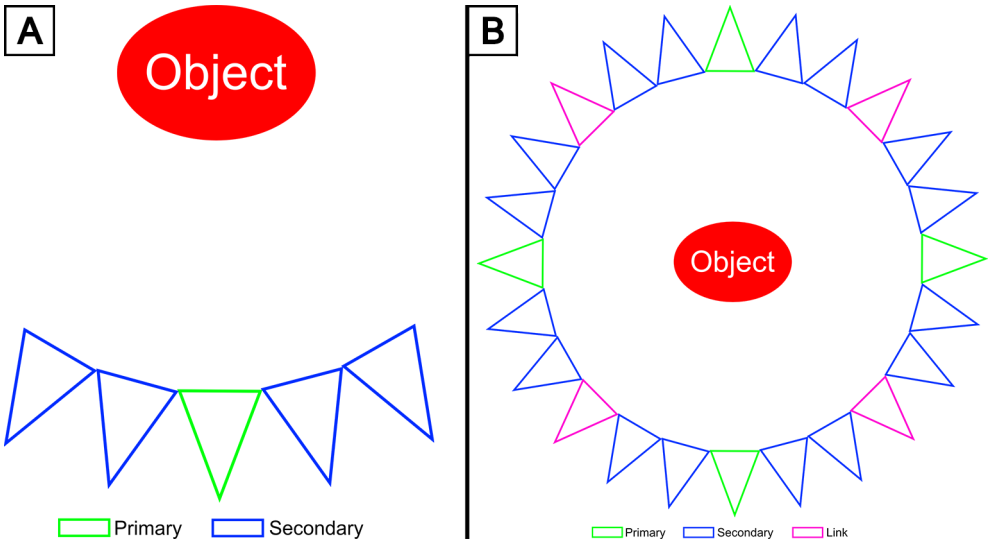

Figure 20: A: Simple convergence method. B: Convergence method with multiple points of view.

3.1.1.1 The gear For this kind of acquisition, fairly common cameras are usually used, from small hand-held compact cameras (or even smartphone cameras) to high end DSLR cameras (see Figure 21). Higher quality camera will produce images with better resolution, lower noise level and sharper details, all things that will improve the end result.

Figure 21: Smartphone – Compact – DSLR

3.1.2 The parallel method

The parallel method is useful when the scene of interest is approximately planar, like a wall, cliff or even part of the Earth's surface (see Figure 22). In that case, none of the pictures will represent the whole scene. On the contrary, every image is a tile of the scene. Images are taken sequentially and each covers parts of the other images around it to ensure that every point of the scene is seen at least twice.

One of the most important thing to consider when planning a survey using the parallel method is the overlap between images. Two kinds of overlap are to be considered: the sequential overlap (also called along-track overlap, see Figure 22) between successive images of the same band (or line) of images and the cross-track overlap, between images of different bands.

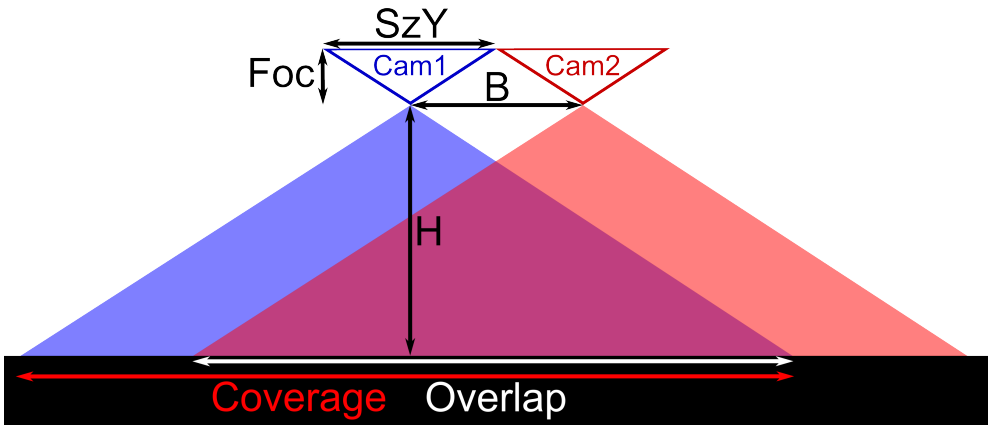

Figure 22: Diagram of the important parameters showing the acquisition of two images along-track.

To ensure that each point is seen at least three times on a single band, the overlap must be over 67%. As a measure of security, and to cope with the actual variations of the terrain/scene, an overlap of 80% is preferred. Cross-track overlap is necessary to link images together, but is less important. It still provides an additional point of view and can help with hidden parts (behind buildings for instance, see Fig. 23 and 24). It is therefore preferred to have a 60% overlap for similar reasons (ensure an actual overlap $> 50\%$). Using a longer focal length creates less hidden part and less foreshortening (see Figure 25).

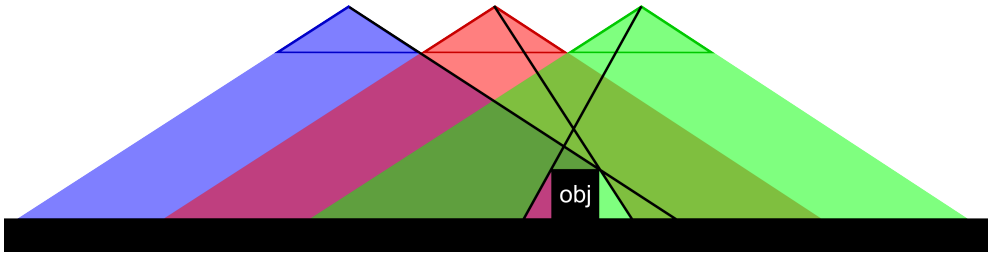

Figure 23: Even with a high overlap, the object in black creates an area that a single image is covering on it's right

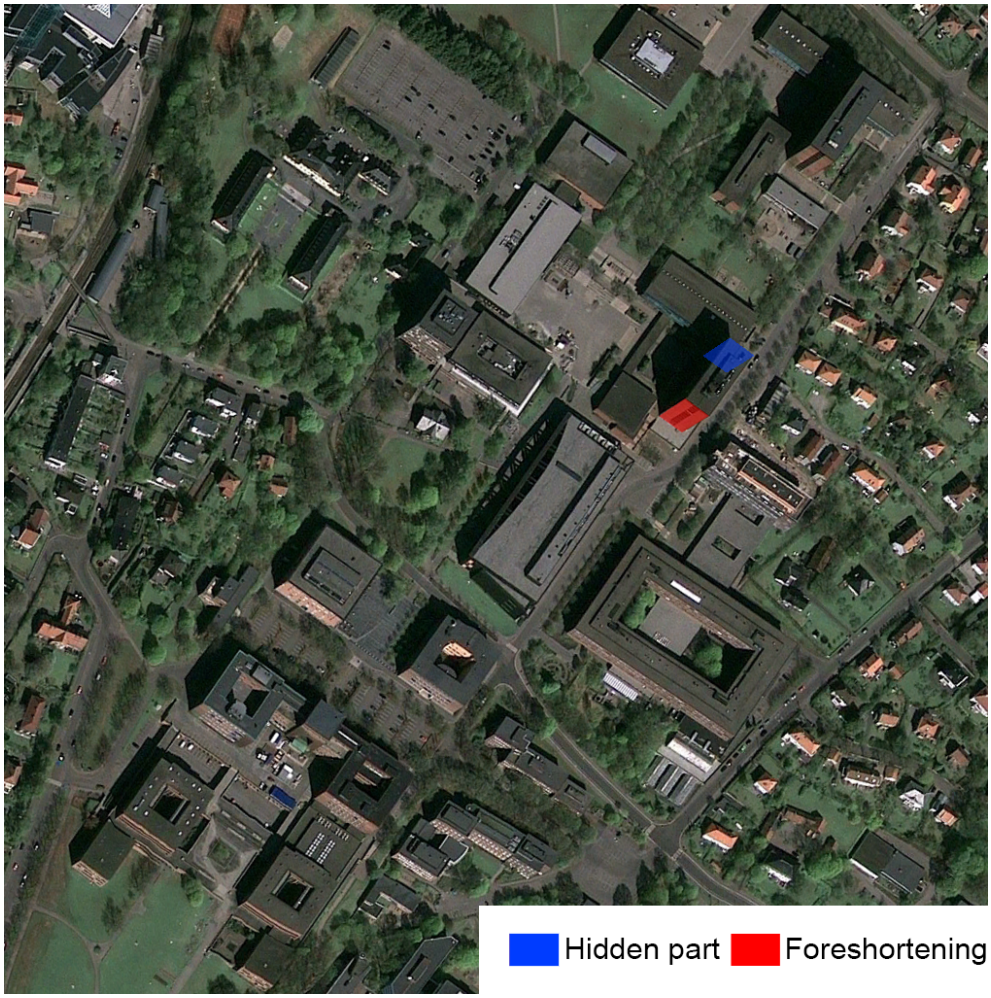

Figure 24: Foreshortening over the university of Oslo from a Google Earth image

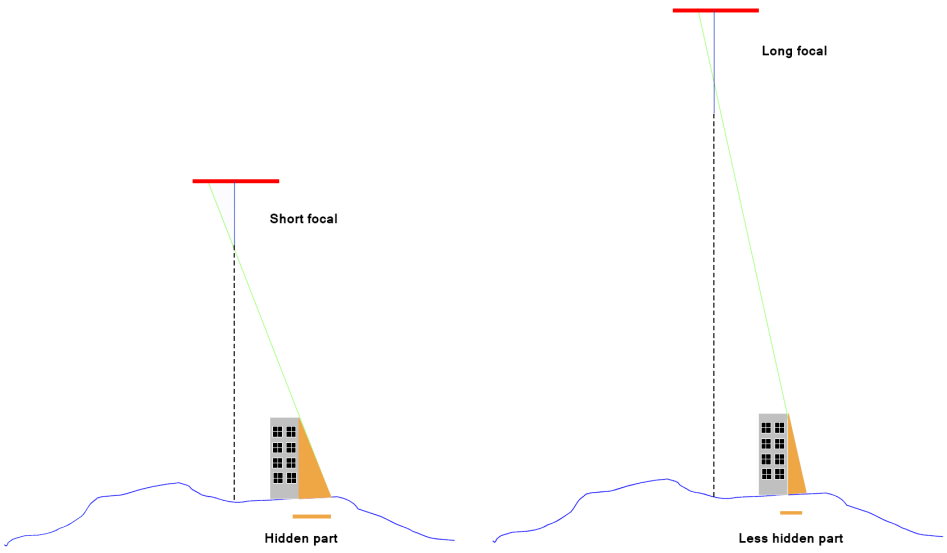

Figure 25: Influence of the focal on hidden parts

3.1.3 Important parameters

For most surveys using the parallel method, the most obvious requirement is that the end product has a given Ground Sampling Distance (GSD, in m), also called ground resolution, the size of a pixel in the output DEM or orthoimage (these might not be the same, it is common to have the orthoimage GSD be half of the DEM GSD). It is a parameter that depends on a multitude of other parameters, and/or will influence the settings of parameters (see also Figure 22):

- The camera's sensor pixel matrix spacing Sz_{Pix} (in mm) ($= Width_{physical}/Width_{pixels}$).
- The camera's focal length **Foc** (in mm).
- The flight height or distance to the scene **H** (in m).

From this, we get the formula:

$$GSD_{image} = \frac{H * Sz_{Pix}}{Foc} \tag{1}$$

Commonly:

$$GSD_{image} \approx GSD_{orthoimage} \tag{2}$$

$$GSD_{orthoimage} = GSD_{DEM}/2 \tag{3}$$

For the parallel method, other parameters are to be taken into account to compute the overlaps (see Figure 26):

- The distance between two consecutive pictures (called Base) **B** (in m).
- The number of pixels in the sensor in the direction of the flight **NbPixY** (in pixels).
- The distance between two lines of acquisition (called cross-track base) **D** (in m).
- The number of pixels in the sensor across the direction of the flight **NbPixX** (in pixels).

The coverages in both directions are then (in m):

$$Cov_{Along-Track} = GSD_{image} * NbPixY \tag{4}$$

$$Cov_{Cross-Track} = GSD_{image} * NbPixX \tag{5}$$

The overlaps are then (in percentage):

$$Overlap_{along-track} = (1 - \frac{B}{Cov_{Along-Track}}) * 100 \tag{6}$$

$$Overlap_{cross-track} = (1 - \frac{D}{Cov_{Cross-Track}}) * 100 \tag{7}$$

B can be computed (or set up) using:

- The velocity of the aircraft **V** (in m/s).
- The frequency of acquisition (time between two pictures) **Freq** (in Hz).

$$B = \frac{V}{Freq} \tag{8}$$

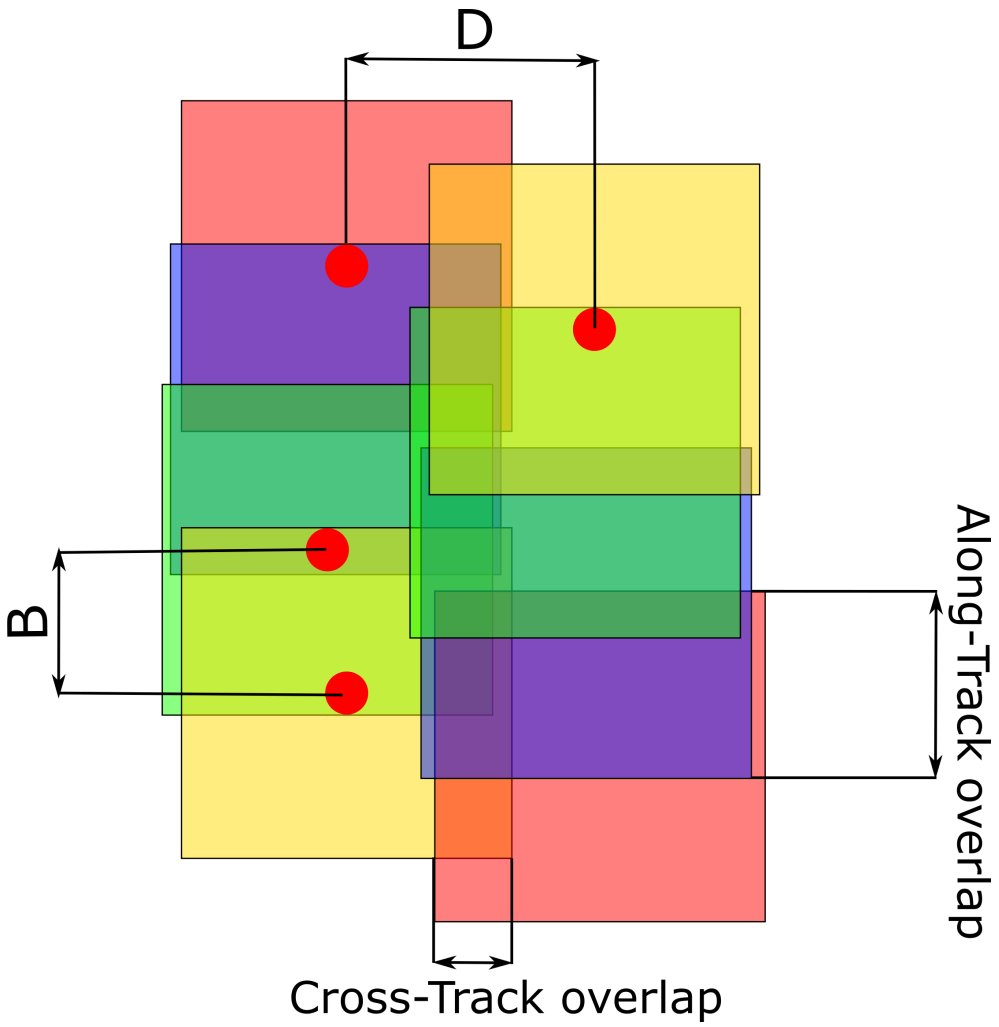

Figure 26: Bases and overlaps

Of course, most parameters can be influenced by choosing different hardware, or configuring it in different ways. However, some other parameters need to be taken into consideration. For instance, the cameras themselves have limits other than their resolution, affecting the amount of light reaching the sensor (aperture **f**, exposition time **ExpTime**) and the sensitivity of the sensor (**ISO**). For a camera embarked on a plane, long exposition times cannot be used because of motion blur. To avoid it, the following condition must be satisfied (with **AcceptableBlur**, the amount of blur considered acceptable in pixels, usually $< 1/2$):

$$V < \frac{GSD_{image}}{ExpTime} * AcceptableBlur \tag{9}$$

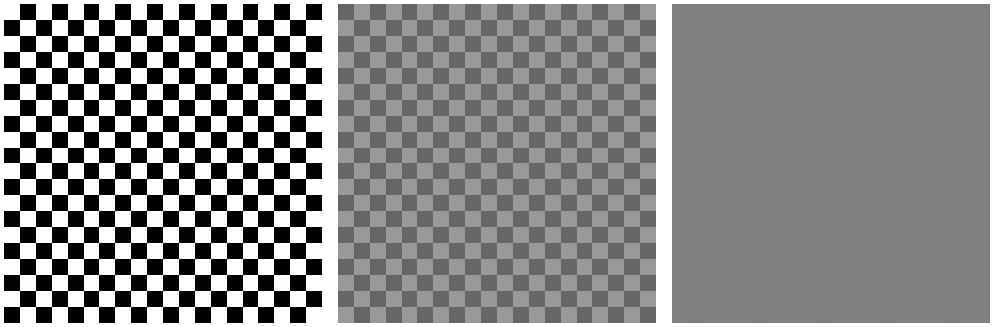

Figure 27: Motion blur on a checker board (0 pixel – 0.5 pixel – 1 pixel)

Figure 28: Example of a 4 pixel motion blur

3.1.3.1 The gear Aerial photography requires the combination of two elements : the camera and the aircraft. Depending on the scale of the project and the precision required, different tools can be used.

The most traditional way is to use a – manned or unmanned – plane (offering a fast and smooth flight) and a specifically designed camera fitted on top of an opening bellow the plane (see Figure 29 and 30). Modern systems also include stabilization platforms, high precision GNSS systems and IMUs (Inertial Measurement Units).

Figure 29: Photogrammetric camera (Vexel Ultracam Xp)

Figure 30: IGN (Institut Geographique National) Beechcraft 200 King Air F-GMGB

The method is however applicable with lighter, smaller and cheaper equipment, such as lighter planes, blimps or helicopters and less specialized cameras.

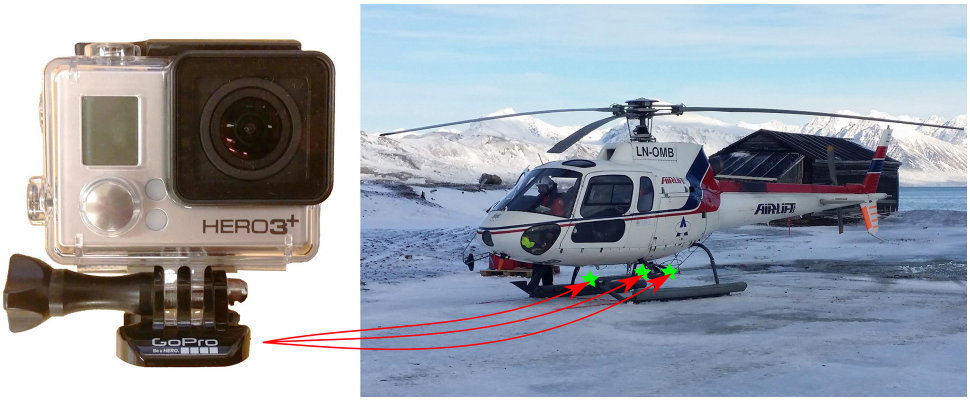

Figure 31: GoPros attached to a hellicopter

Figure 32: Blimp equipped with a camera

Figure 33: Small UAV (DJI Mavic 2 Pro)

3.2 Tie points

To be able to use several images in a set, it is necessary to know how they are related. To estimate that, tie points (TP) are required.

3.2.1 Historical tie points

Historically, tie points where gatherd manually, 6 per image couple (close to the minimum for relative orientation, see section 3.7.2), at the Von Gruber positions (see Figure 34). In the 90s, automatic algorithms could look for tie points in the areas around optimal Von Guber points.

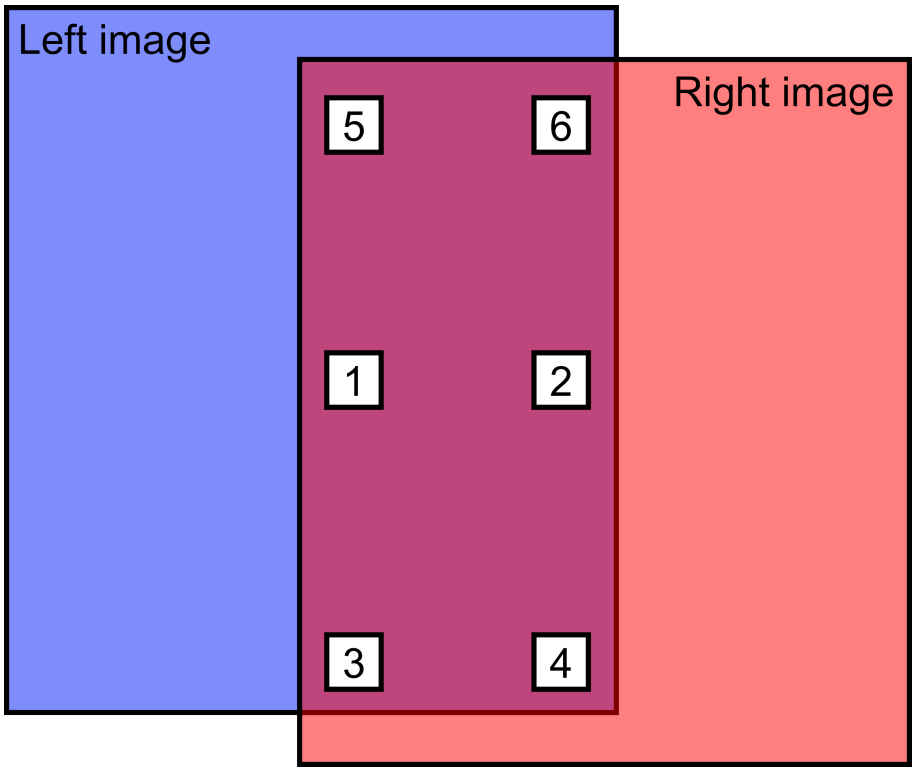

Figure 34: Von Gruber positions for tie points

3.2.2 Modern tie points

Exponentially increasing computing power and the development of automatic tie point detection algorithms such as SIFT [Lowe, 2004] (Scale Invariant Feature Transform, one of the first robust multipurpose automatic tie point extraction and description algorithm), SURF [Bay et al., 2008], ASIFT [Morel and Yu, 2009] and others allowed for the automatic search of tie points in unorganized set of images. The number of tie points that are then available is orders of magnitude higher and potential outliers (mismatch) can be filtered out. Figure 35 shows the tie points for an image pair and a group of obvious outliers.

The process of automatic tie point collection is divided in three steps:

- The identification of remarkable points. The remarkable quality is defined differently by different algorithm. SIFT searches points at different sub-resolution of an image that are either brighter or darker that all other adjacent points.

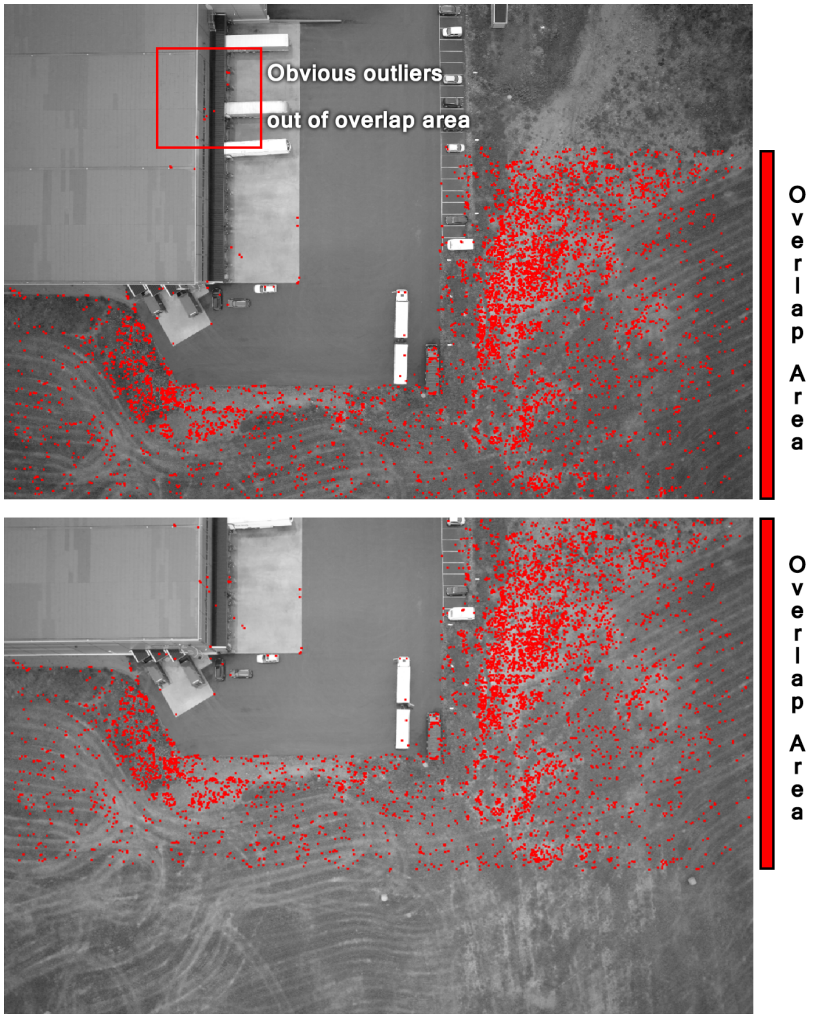

Figure 35: Tie Points detected with SIFT on a pair of images from a drone.

- Assigning a descriptor to these points. The descriptor is usually given in a vector space quite different from the image grid of value. SIFT creates a normalized descriptor of the keypoint using the gradients of the area, rotated so the strongest gradient is oriented "North" (see Figure 36).
- The descriptors of points from different images are matched together to define the tie points.

The descriptors from SIFT, SURF and ASIFT are completely robust to rotation, global brightness change and scaling, and proved to be reliable even with noisy images or difference in viewing angle over 30 degrees, even if such high angles are to be avoided (recommended value is maxiumum 15 degrees between optical rays).

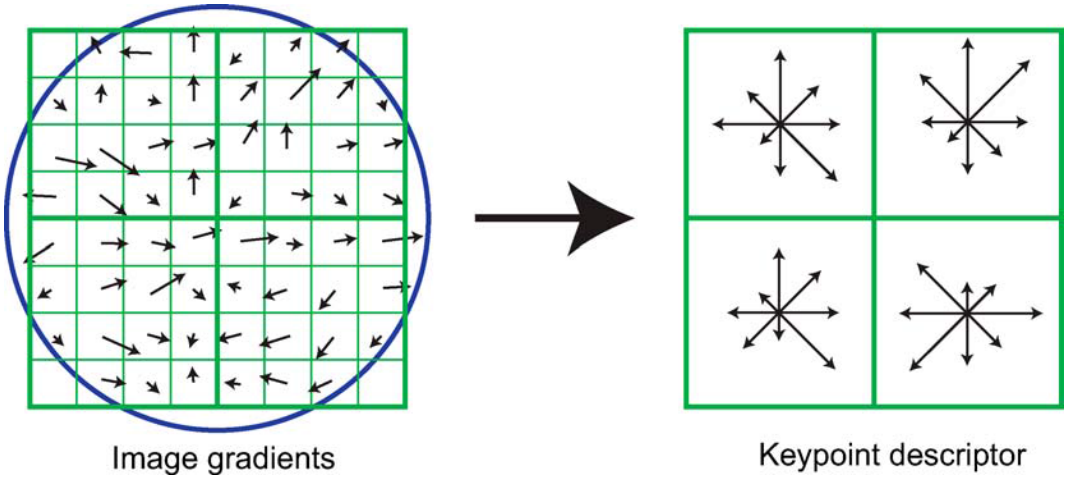

Figure 36: The SIFT keypoint descriptor

3.3 A few notions of function fitting

In the next section, we will have to fit models to observations. To do so, we always go through the same steps :

- Gather observations.
- Choosing a model that could fit them.
- Make sure the model if valid for unobserved points.

To model complicated phenomenons, simple models often don't offer a good fit. It is tempting to try using models with a lot of parameters (high degree polynomials...) that can fit very well the observations, but over-parametrization can lead to over fitting. The Figure 37 shows the effect of too poor and too rich models.

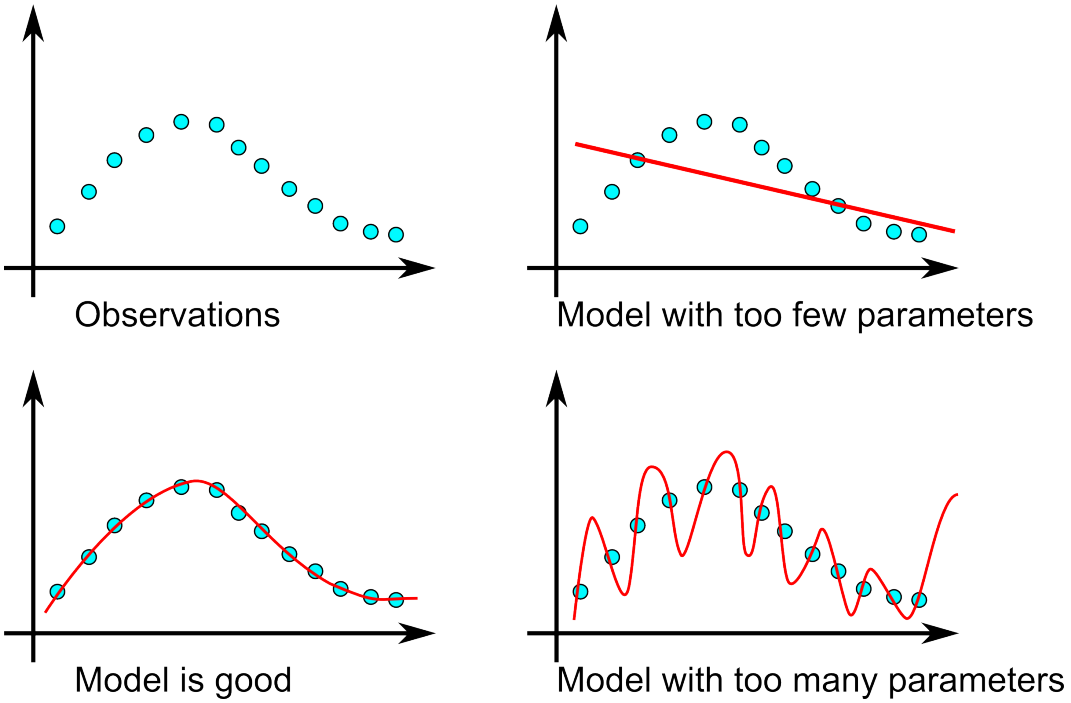

Figure 37: Different models fitting observations

A few rules of thumb:

- If the phenomenon have know physical cause, it is best to base the model on them.
- Over parametrization is more costly in computing time.
- If the number of observations is sensibly higher than the number of parameters, it is often best to over parametrize.
- Extrapolating is a dangerous bet (see Figure 38).

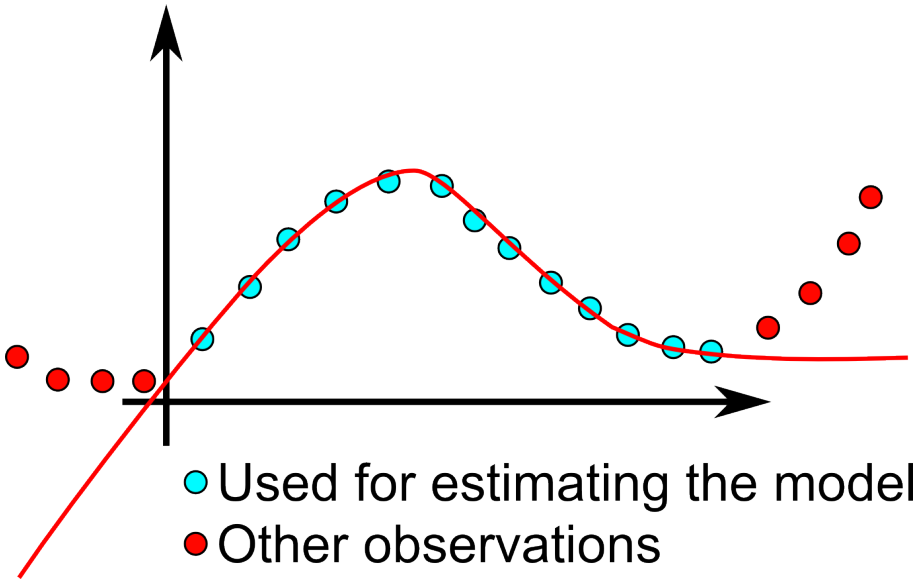

Figure 38: Model is wrong outside of the observations' zone

In case of observations presenting a large number of outliers, the RANSAC (RANdom SAmple Consensus) method should be used :

1. Select a random set of observations big enough to estimate the parameters of the model.

2. Count the number of observations agreeing with the model (fitting error $<$ threshold).

3. Reiterate steps 1 and 2.

4. Select the run with the highest number of agreeing points (see Figure 39).

5. Refine the chosen parameters using all the agreeing points.

Remark : The method can also be used to find several objects in a set of observations (if distinct sets of observations yield good scores).

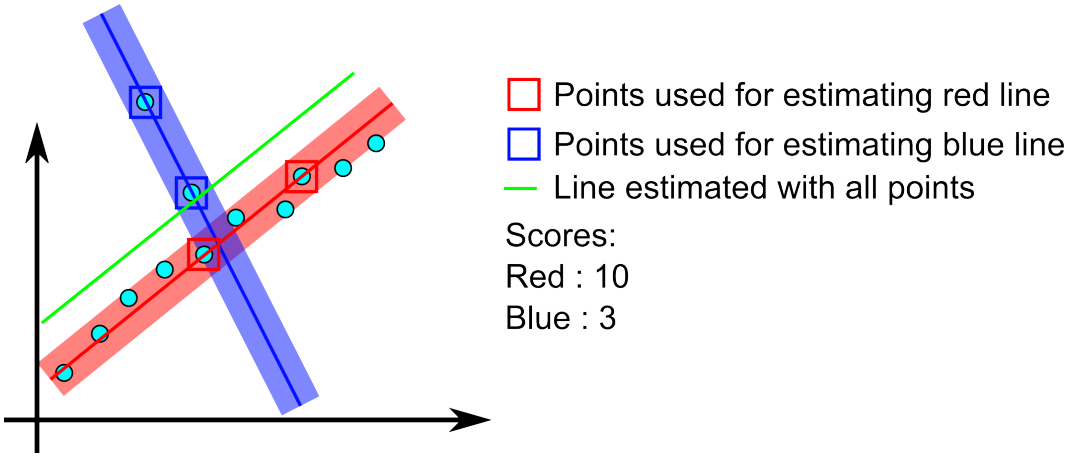

Figure 39: RANSAC method on estimating a line

3.4 Short overview of 2D interpolation

Wikipedia : Interpolation is a method of constructing new data points within the range of a discrete set of known data points.

In image processing, interpolation is used to compute the color value of a point the is not exactly the center of a pixel, but an intermediate value. It is usefull when resizing or rotating images, or when performing orthorectification (see Section 3.10).

A lot of interpolation methods exists, they go from very simple like Nearest Neighbor Interpolation up to very complex method taking into account contextual information. Bellow is an overview of the simplest methods.

3.5 Nearest Neighbor Interpolation

As the name suggest, Nearest Neighbor Interpolation simply looks up the value of the data point geometrically closest to the query point.

3.6 Linear/Bilinear Interpolation

Linear interpolation (one dimensional) takes the weighted average of the two neighboring points. Bilinear Interpolation is the two-dimensional extension of that concept.

Suppose that we want to find the value of the unknown function f at the point (x, y). It is assumed that we know the value of f at the four points $Q_{11} = (x_1, y_1)$, $Q_{12} = (x_1, y_2)$, $Q_{21} = (x_2, y_1)$, and $Q_{22} = (x_2, y_2)$ like in Figure 40.

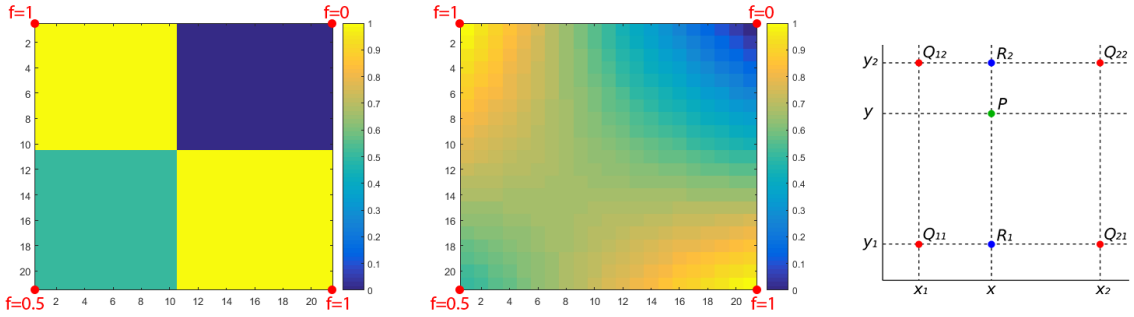

Figure 40: Left : Nearest Neighbor Interpolation of the four corners in a 21x21 grid. Middle : Bilinear Interpolation. Right : Schema for Bilinear Interpolation

We first compute the value for the points R_1 and R_2, weighted averages of the values in respectively Q_{11} and Q_{21}, and of Q_{12} and Q_{22} :

$$f(x, y_1) = f(R_1) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x - x_1}{x_2 - x_1} f(Q_{21}) \tag{10}$$

$$f(x, y_2) = f(R_2) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x - x_1}{x_2 - x_1} f(Q_{22}) \tag{11}$$

Then we can compute the weighted average of the values of R_1 and R_2 for the query point P :

$$f(x, y) = f(P) \approx \frac{y_2 - y}{y_2 - y_1} f(R_1) + \frac{y - y_1}{y_2 - y_1} f(R_2) \tag{12}$$

As Figure 40 clearly shows, this method gives much softer transitions between points. Better, more complicated methods exist and give even better results.

3.7 Camera calibration and orientation

Once the images are acquired, they need to be oriented (also called aerotriangulated), that is put together in a unique geometric coordinate system, and the camera(s) need to be calibrated. Using TPs, Both these operations can be done independently or together through the structure-from-motion (SfM) method [Snavely, 2010].

The orientation of an image for perspective cameras (function $\mathcal{O}$) can be described by the general equation 13. It describes the transformations required to convert the coordinates of a point in a given 3D coordinate system (called the Relative Space, *rs*, in Eq. 13) into its coordinate in one in the images.

$$\begin{pmatrix} i_{L,K} \\ j_{L,K} \end{pmatrix} = \mathcal{O}_K \begin{pmatrix} {}^{rs}x_L \\ {}^{rs}y_L \\ {}^{rs}z_L \end{pmatrix} = \mathfrak{I}\left(\pi\left(R_K * \left(\begin{pmatrix} {}^{rs}x_L \\ {}^{rs}y_L \\ {}^{rs}z_L \end{pmatrix} - C_K\right)\right)\right) \tag{13}$$

Where:

- L is an object.
- K is an image.
- $({}^{rs}x_L; {}^{rs}y_L; {}^{rs}z_L)$ are the coordinates of the object L in Relative space coordinates.
- C_K is the coordinates of the optical center of the camera K in Relative space coordinates.
- R_K is the rotation matrix from Relative space coordinates to Camera coordinates.
- π is the function projecting points in Camera coordinates to a canonical 2D space.
- $\mathfrak{I}$ is the function of the camera parameters (Focal, distortions, sensor size (in mm and pixels)...) converting canonically projected points into Image coordinates.
- $(i_{L,K}, j_{L,K})$ are the pixel coordinate of the object L in image K.

For satellite pushbroom sensors (also called digital scanners, see Section 1.2.6), an other formulation is necessary. Formulations similar in spirit with the one for perspective cameras exist and associate each line of the image to an individual function, but the solution preferred in modern software and by satellite data providers is the Rational Polynomial Coefficient functions – RPC –, also called Rational Function Models – RFM – [Tao and Hu, 2001]. The direct RPC computes the transformation from image to geographical coordinates (see Equations (14), (15) and (18) and the inverse RPC computes the transformation from geographical to image coordinates (see Equations (16)–(18); they are rational function polynomial equations of the normalized image and geographical coordinates (scaled to a unit cube), defined as:

$$Lon_{norm} = \frac{P_1(Col_{norm}, Row_{norm}, h_{norm})}{P_2(Col_{norm}, Row_{norm}, h_{norm})} \tag{14}$$

$$Lat_{norm} = \frac{P_3(Col_{norm}, Row_{norm}, h_{norm})}{P_4(Col_{norm}, Row_{norm}, h_{norm})} \tag{15}$$

$$Col_{norm} = \frac{P_5(Lon_{norm}, Lat_{norm}, h_{norm})}{P_6(Lon_{norm}, Lat_{norm}, h_{norm})} \tag{16}$$

$$Row_{norm} = \frac{P_7(Lon_{norm}, Lat_{norm}, h_{norm})}{P_8(Lon_{norm}, Lat_{norm}, h_{norm})} \tag{17}$$

with:

$$\begin{aligned} P_i(X,Y,Z) = C_1 + C_2X + C_3Y + C_4Z + C_5XY + C_6XZ + C_7YZ + C_8X^2 + C_9Y^2 + C_{10}Z^2 \\ +C_{11}XYZ + C_{12}X^3 + C_{13}XY^2 + C_{14}XZ^2 + C_{15}X^2Y \\ +C_{16}Y^3 + C_{17}YZ^2 + C_{18}X^2Z + C_{19}Y^2Z + C_{20}Z^3 \end{aligned} \tag{18}$$

3.7.1 Defining coordinate systems

Several coordinate systems are to be considered here.

3.7.1.1 Image coordinates (i_L, j_L)

These are the coordinates of a point in the image. The unit is the pixels, the x axis is the row of the pixel and y is the line. The origin is the top left corner.

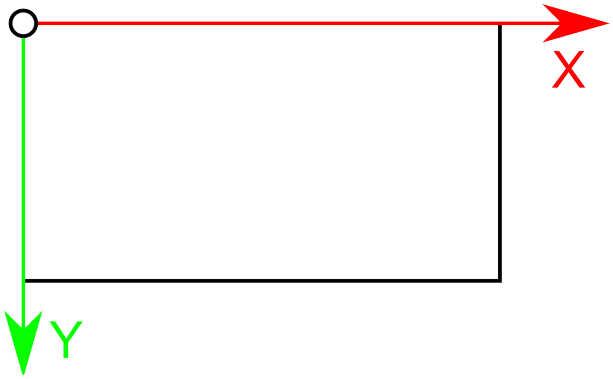

Figure 41: Image coordinates system

3.7.1.2 Canonical 2D space $({}^sx_L; {}^sy_L)$

These are the coordinate of a point projected in a 2D space. The units are meters (millimeters in fact since sensors are small), the x axis is to the right and y axis to the top. The origin is the orthogonal projection of the optical center.

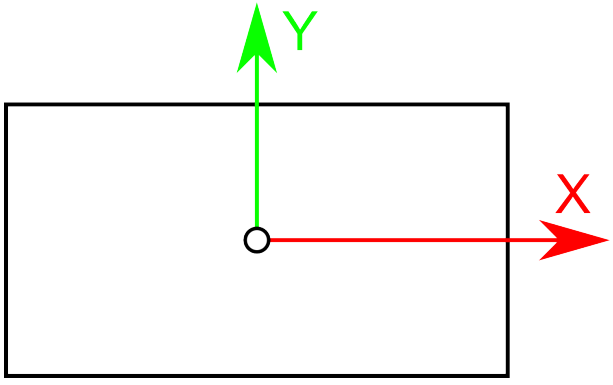

Figure 42: Canonical 2D space coordinates

3.7.1.3 Camera coordinates $({}^cx_L; {}^cy_L; {}^cz_L)$

These are the coordinates of the point in a system where the optical center of the camera is at $\begin{pmatrix}0\\0\\0\end{pmatrix}$ and the camera is looking down the Z axis. The units are meters.

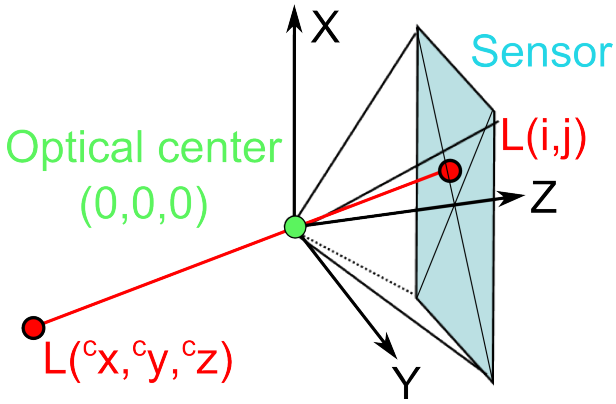

Figure 43: Camera coordinates

3.7.1.4 Relative space coordinates $({}^{rs}x_L; {}^{rs}y_L; {}^{rs}z_L)$

These are the coordinates of the point in the system in which the cameras are oriented together. Usually, it is the same as the camera coordinates of one of the camera.

3.7.1.5 World coordinates $({}^{w}x_L; {}^{w}y_L; {}^{w}z_L)$

These are the coordinates of the point in "real world" coordinates. If the system is geolocalized, it is the actual coordinate of the point. Of course, units are meters.

3.7.2 R_K and C_K : Relative camera orientation

We first want to solve for the position (C_K) and looking direction (called orientation, (R_K)) of the camera.

$$\begin{pmatrix}{}^{c}x_L\\{}^{c}y_L\\{}^{c}z_L\end{pmatrix} = R_K * \left(\begin{pmatrix}{}^{rs}x_L\\{}^{rs}y_L\\{}^{rs}z_L\end{pmatrix} - C_K \right) \tag{19}$$

3.7.2.1 Note :

In the algorithms presented here, a number of points are required. We usually have a lot more points available and we can use them in a RANSAC procedure to ensure robustness to outliers and noise.

3.7.2.2 Step 1 : Two cameras

Relative orientation is the method aimed at knowing the position and orientation of an image relatively to an other. In other words, finding $({}^{rs}R_{c_1}, {}^{rs}R_{c_2})$ and the positions of both cameras $({}^{rs}PC_1, {}^{rs}PC_2)$, with rs being the relative space. These are 12 unknown (2*3 rotations and 2*(x,y,z)). To give a base reference to the system, we consider the relative space to be the first camera's space, so :

$${}^{rs}R_{c_1} = \begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix} \quad {}^{rs}PC_1 = \begin{pmatrix}0\\0\\0\end{pmatrix} \tag{20}$$

Also, in order to fix the scale :

$$x_1 - x_2 = 1 \tag{21}$$

In other words, we just need to find ${}^{c_1}R_{c_2} = {}^{rs}R_{c_2}$ and two elements of $\overrightarrow{PC_1PC_2}$ now. These are a total of 5 parameters (3 rotations and 2 translation). In theory, we then need at least five tie points to solve the system. Algorithms exist to solve this problem, but they are out of the scope of this course.

3.7.2.3 Step 2 : More cameras

Once the first two cameras are oriented, we just need to link the others. Two methods are possible :

- Keep using the two cameras method incrementally.
- Use the positions of the points used for other images (computed as a byproduct) and the resection algorithm.

3.7.3 π : from Camera to Canonical coordinates

We now have the coordinates of points in camera coordinates. We need to project them into a 2D space through a canonical projection (the function π).

With $\begin{pmatrix} {}^cx_L \\ {}^cy_L \\ {}^cz_L \end{pmatrix}$ the position (in m) of a point L in camera coordinates, we can write the following equation, where $\begin{pmatrix} {}^sx_L \\ {}^sy_L \end{pmatrix}$ would be the position of a point canonically projected through a perfect optic of $Foc = 1$:

$$\begin{pmatrix} {}^sx_L \\ {}^sy_L \end{pmatrix} = \pi\left(\begin{pmatrix} {}^cx_L \\ {}^cy_L \\ {}^cz_L \end{pmatrix} \right) \tag{22}$$

Then, simply using the Thales theorem, on the system shown in Figure 44 we have the formula :

$$\begin{pmatrix} {}^sx_L \\ {}^sy_L \end{pmatrix} = \pi\left(\begin{pmatrix} {}^cx_L \\ {}^cy_L \\ {}^cz_L \end{pmatrix} \right) = \begin{pmatrix} {}^cx_L/{}^cz_L \\ {}^cy_L/{}^cz_L \end{pmatrix} \tag{23}$$

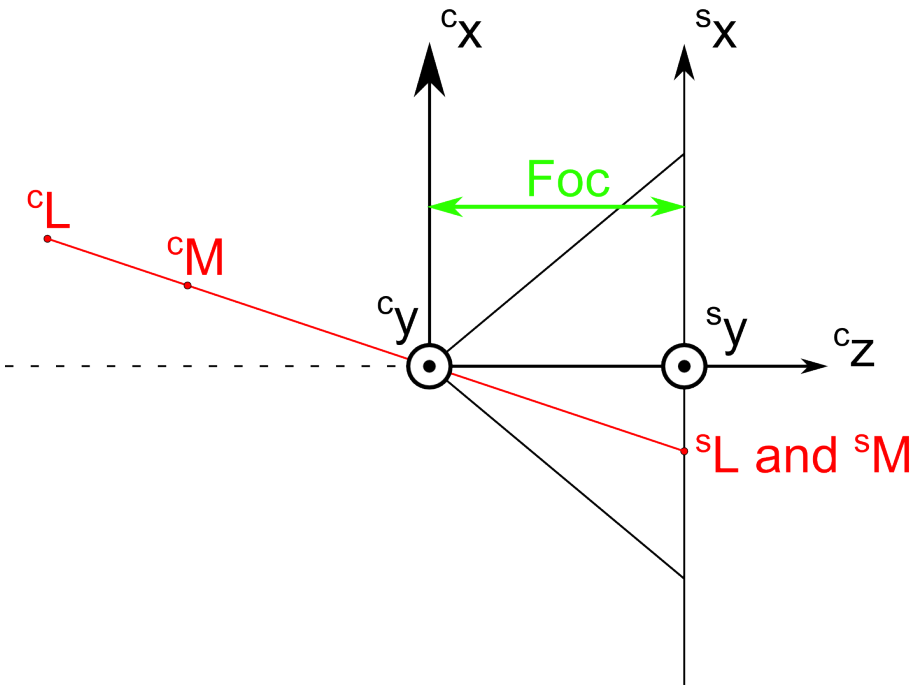

Figure 44: Projection from camera coordinates to canonical coordinates

We can notice in Figure 44 that the same point in canonical coordinates can be obtained from a whole line in camera coordinates. It is therefor impossible to have a function π^{-1} that would provide the 3D coordinate back. We can however give back the parameters of the line : its direction is $\begin{pmatrix} {}^s x_L \\ {}^s y_L \\ 1 \end{pmatrix} = \begin{pmatrix} {}^c x_L / {}^c z_L \\ {}^c y_L / {}^c z_L \\ 1 \end{pmatrix}$ and its origin is $\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$.

3.7.4 $\mathfrak{I}$: Camera calibration

For this part, we want to solve the following equation by finding the camera internal parameters :

$$\begin{pmatrix} i_L \\ j_L \end{pmatrix} = \mathfrak{I} \begin{pmatrix} {}^s x_L \\ -{}^s y_L \end{pmatrix} \tag{24}$$

The search for the parameters is initialized by the meta data of the image (or manually provided) : size of the sensor (in pixels and mm) and focal.

3.7.4.1 Parameters and general formulas

3.7.4.2 A - Focal length and pixel size

We first need to scale, orient and center the system properly, using the focal (Foc, in mm) and the spacing between two pixels on the sensor (Sz_{Pix} in mm/pix), inverting the y axis, and moving the origin by the coordinates of the principal point (PP, in pixels, considered equal to the center of the image). For an ideal camera we would have :

$$\begin{pmatrix} {}^{id} i_L \\ {}^{id} j_L \end{pmatrix} = \begin{pmatrix} i_{PP} \\ j_{PP} \end{pmatrix} + Foc * \begin{pmatrix} {}^s x_L \\ -{}^s y_L \end{pmatrix} / Sz_{Pix} \tag{25}$$

3.7.4.3 B - Geometrical distortion

We now face the optical imperfection of lenses. Real lenses can't provide perfect Gaussian conditions for the optics (lenses are not infinitely thin) and therefor introduce distortion. Then general form of $\mathfrak{I}$ is then :

$$\begin{pmatrix} i_L \\ j_L \end{pmatrix} = D\left(\begin{pmatrix} i_{PP} \\ j_{PP} \end{pmatrix} + Foc * \begin{pmatrix} {}^{s}x_L \\ -{}^{s}y_L \end{pmatrix} / Sz_{Pix} \right) \tag{26}$$

Or :

$$\begin{pmatrix} i_L \\ j_L \end{pmatrix} = D\left(\begin{pmatrix} {}^{id}i_L \\ {}^{id}j_L \end{pmatrix} \right) \tag{27}$$

With D being a more or less complex function of $\begin{pmatrix} {}^{id}i_L \\ {}^{id}j_L \end{pmatrix}$, the "ideal camera image coordinates".

3.7.4.4 Distortion models

The distortion can me modeled in different ways. We know however that the phenomenon is the result of the projection of a cylindrically symmetric optical system (see Figure 45), and therefore close to be radial.

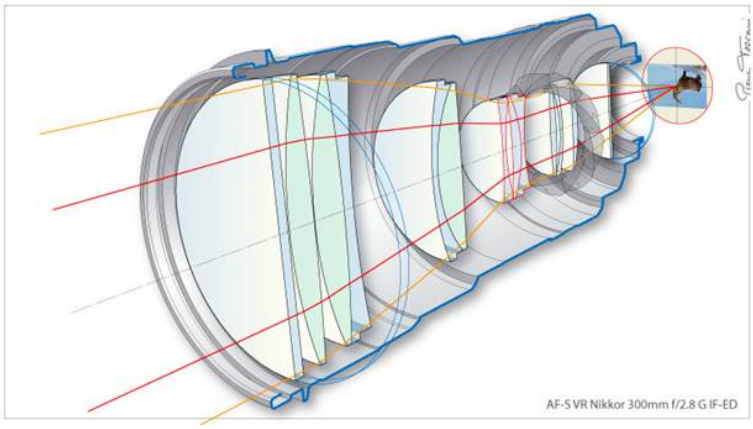

Figure 45: Schematics of a lens

The distortion is then a function defined by :

- a distortion center **C**.
- a radial distortion function $D_r(d)$, with d the distance to **C**.

We also know that lenses have extremely regular surfaces, so we can hypothesize that D_r is C^∞ (smooth). Therefore we can model it with an odd polynomial function :

$$D_r(d) = d + \alpha * d^3 + \beta * d^5 + ... \tag{28}$$

We then have :

$$D(L) = C + \overrightarrow{CL}(1 + \alpha * d^2 + \beta * d^4 + \gamma * d^6) \tag{29}$$

But more complex models are often used to account for misalignment of the lenses or their imperfection (lower order correction). For example Brown's model :

$$\begin{pmatrix} i_L \\ j_L \end{pmatrix} = \begin{pmatrix} {}^{id}i_L * (1 + K_1 * d^2 + K_2 d^4 + ...) + (P_2(d^2 + 2 * {}^{id}i_L^2) + 2P_1 {}^{id}i_L {}^{id}j_L) * (1 + P_3 d^2 + P_4 d^4 + ...) \\ {}^{id}j_L * (1 + K_1 * d^2 + K_2 d^4 + ...) + (P_1(d^2 + 2 * {}^{id}j_L^2) + 2P_2 {}^{id}i_L {}^{id}j_L) * (1 + P_3 d^2 + P_4 d^4 + ...) \end{pmatrix} \tag{30}$$

The parameters are :

- d the distance between ideal point and the distortion center.
- K_n the radial distortion coefficients.
- P_n the tangential distortion coefficients.

Or even the Ebner model (explained in Figure 46).

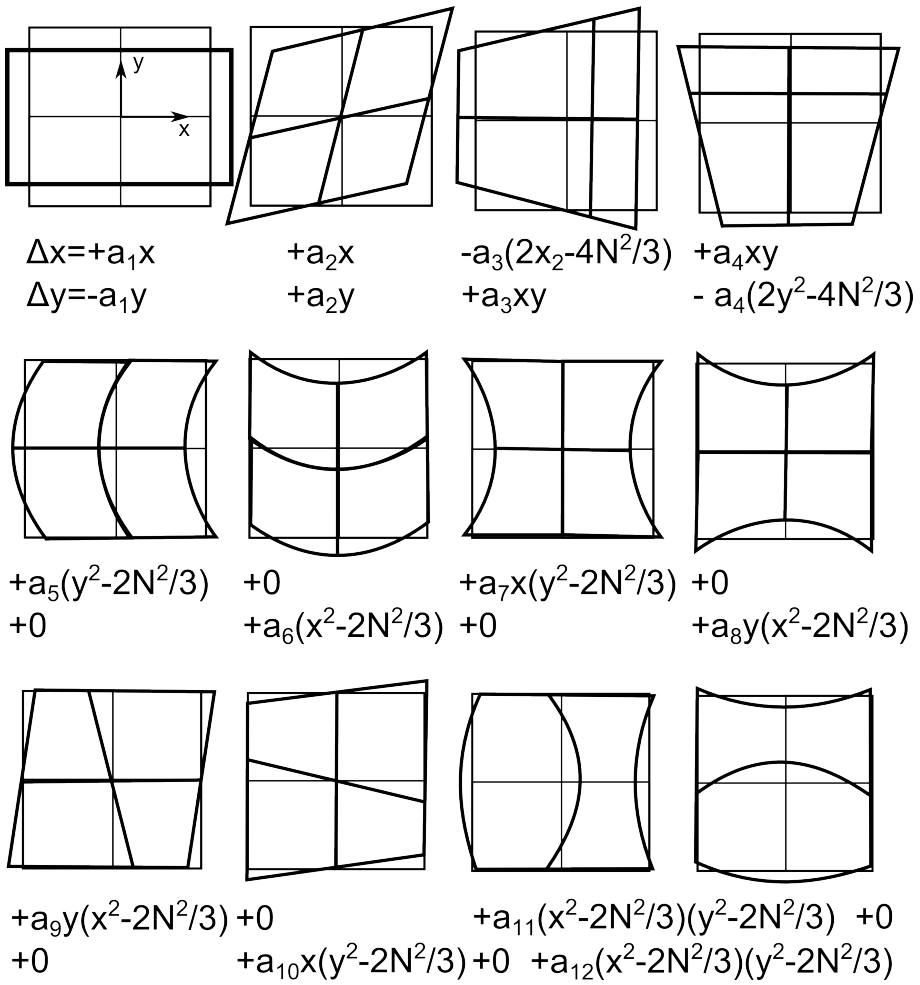

Figure 46: Ebner distortion model (a_i parameters , N the largest image coordinate)

3.7.4.5 Estimating calibration parameters To estimate the calibration parameters, calibration images must be taken. A group of images with tie points between them are given with an external orientation and an internal orientation model has to be fitted on the observations.

Given the non linearity of the models and the number of parameters, initialization values are needed. For a "normal" lens, they are usually :

- Function D is the identity (no distortion).
- PP is the center of the image.
- Approximate Foc is given by the user (or image meta-data).

Then fitting algorithms are used while the parameters are released, starting with the most influential ones (usually Foc first).

(a) Original image with high distortion

(b) Image corrected for distortion

3.8 Georeferencing

Once all the cameras are oriented relatively to each other, we almost always want an "absolute" referencing (either to a cartographic system, or to a local system for scale). To be able to georeference a DEM, orthoimage and/or 3D model through absolute orientation, some geodetic information about the scene is required. This can be achieved through Ground Control Points (GCP), points visible in the imagery with known coordinates in the desired coordinate system, either collected through GNSS surveying or other topographic surveying method, or by using other georeferenced products (previously generated DEM and orthoimage, or a map for example). The minimum mathematical requirements to achieve georeferencing are:

- Find at least 3 non-collinear points that are seen in at least 2 images (optical ray intersection will give their position in relative space coordinates).
- Know the position of these points in the targeted, "Absolute/World" coordinate system.
- Find the 7 parameters of the transformation between the two systems (3 rotations (${}^{w}R_{rs}$), 3 translations (${}^{w}C_{rs}$) and scaling (λ) – see Equation 31).

$$ {}^{w}L = \lambda * {}^{w}R_{rs} * ({}^{rs}L - {}^{w}C_{rs}) \tag{31} $$

For the convergent method, reference points should uniformly surround the scene and be visible in the images.

For the parallel method (e.g. for typical aerial survey), the equipment in reference points should be as follows:

- Surrounding the area of interest (zones outside of the reference polygon rely on extrapolation).

- Some points inside of the area of interest (to avoid a "dome" or "banana" effect [James and Robson, 2014]).
- It is best to have XYZ for every point, but some points can be Z only or XY only.
- Having more points than strictly necessary is always preferable, and unused GCPs can be used as check points to evaluate the precision and accuracy of the data.

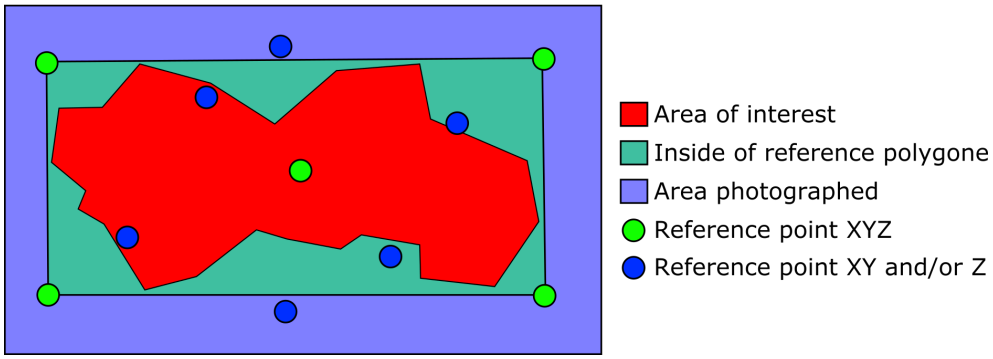

Figure 47: Position of the minimal amount of reference points for an aerial survey.

An other way to obtain information on the position (and possibly orientation) of the camera for each pictures is through camera- or system-integrated GNSS (and possibly an inertial measurement unit – IMU) systems. This information can be used as a first approximation, as complementary data or, if of sufficient quality, on its own.

3.9 Dense correlation

Once the position, orientation and camera parameters of each image are known, the final reconstruction of the 3D information can start. This process, called dense correlation or dense multiview stereopsis [Furukawa and Ponce, 2010], exploits image correlation to compute the geometry of the scene. Using the internal and external orientation of cameras, it is possible to project points in the 3D space in the photographed scene back into the images, or a pixel in an image into an optical ray in the 3D scene. This allows for the search of homologous points in several images (see Fig. 48). Strategies to find these homologous points are described in Sections 3.9.2 and 3.9.3.

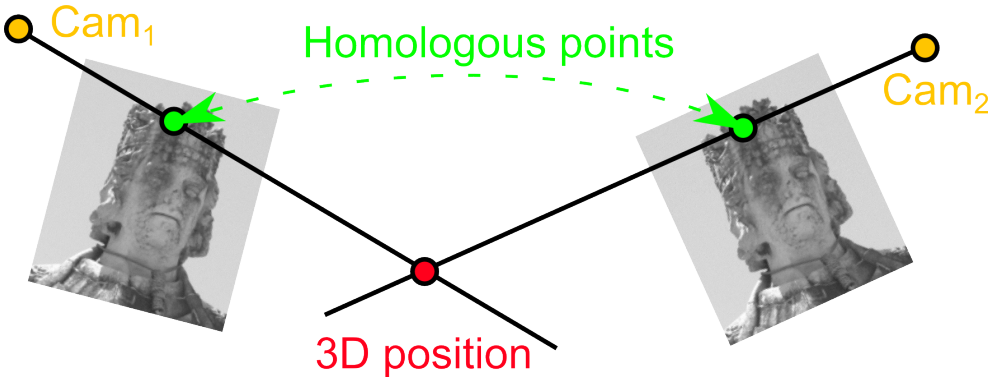

Figure 48: Intersecting optical rays coming from homologous points in two images.

3.9.1 Image correlation

The principle of image correlation is to find a common template in several images by scoring the resemblance between the different excerpts (see Fig. 49). A match is identified by the best score. It is best to look for normalized patterns so that a global brightness change will not affect the results. The following formula gives the normalized cross correlation score between a k-by-k correlation window (with $k = 2*n+1$) of an image and a template of the same size:

$$Corr(x,y) = \frac{\sum_{u=-n}^{n}\sum_{v=-n}^{n}(f(x+u,y+v)-\overline{f}_{x,y})(t(u,v)-\overline{t})}{\sqrt{\sum_{u=-n}^{n}\sum_{v=-n}^{n}(f(x+u,y+v)-\overline{f}_{x,y})^2 \sum_{u=-n}^{n}\sum_{v=-n}^{n}(t(u,v)-\overline{t})^2}} \quad (32)$$

Where:

- t is the template from the master image (its center is $t(0,0)$).
- $\overline{t}$ is the mean of the template.
- f is the slave image.
- $\overline{f}_{x,y}$ is the mean of f in the region $f(x \pm n, y \pm n)$.

Note that this function is not applicable to 1-by-1 correlation window or to constant templates (this case would result to a 0/0 result). This is usually not an issue because:

- A 1-by-1 correlation window would not give useful information (for example, in an 8 bits grey scale image, only 256 different "windows" would be possible).
- A constant template does not have any feature to help correlation, so correlation is set to 0 in that case.

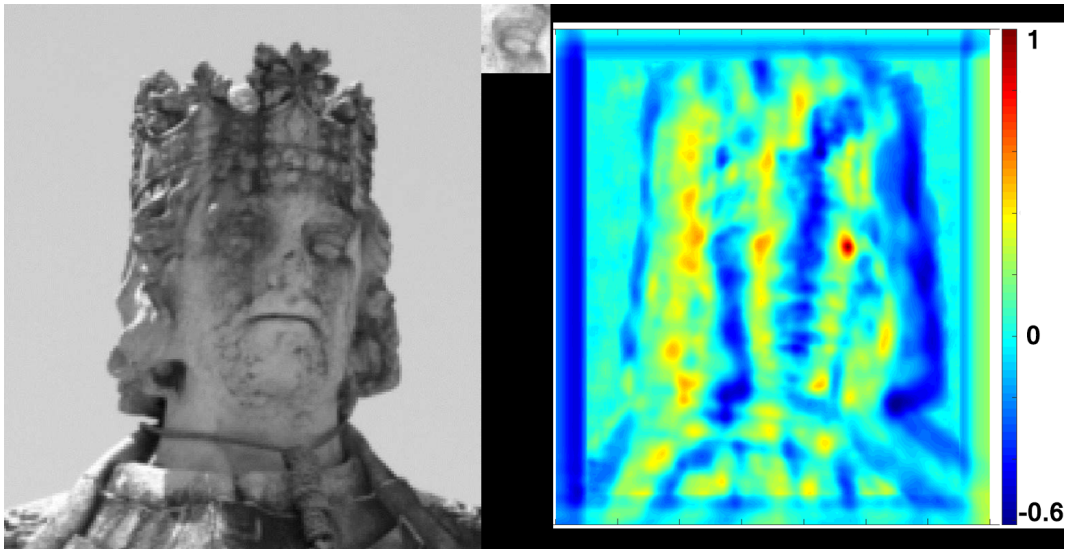

Figure 49: Left : Image to be searched; Middle : Template to look for; Right : Matlab normxcorr2

3.9.1.1 Influence of the correlation window size In the function 32, we use correlation windows. They are the size of the template from an image we are trying to find in another image, and their size can be influential on the result.

A small window :

- Give fine details and accurately reconstruct high frequencies in the model.
- Is fast to compare.
- Introduces high noise since the simplicity of the template means that higher scoring templates can be found away from the "real" solution.

A big window :

- Lose fine details in the high frequencies of the scene.
- Is slow to compare.
- Produces few outliers and a generally more coherent model.

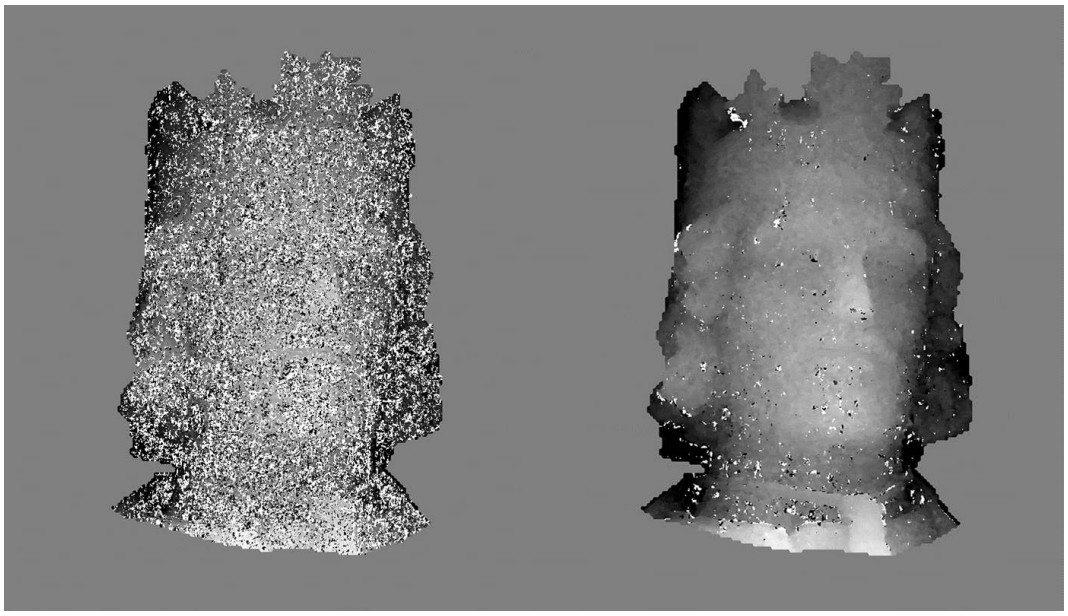

Figure 50: Left : 3-by-3 ; Right : 9-by-9

3.9.2 Convergent method

For the convergent method (see Fig. 51), images are grouped into subsets that cover the same zone of the scene, and the image at the center of the group is defined as the *master image*.

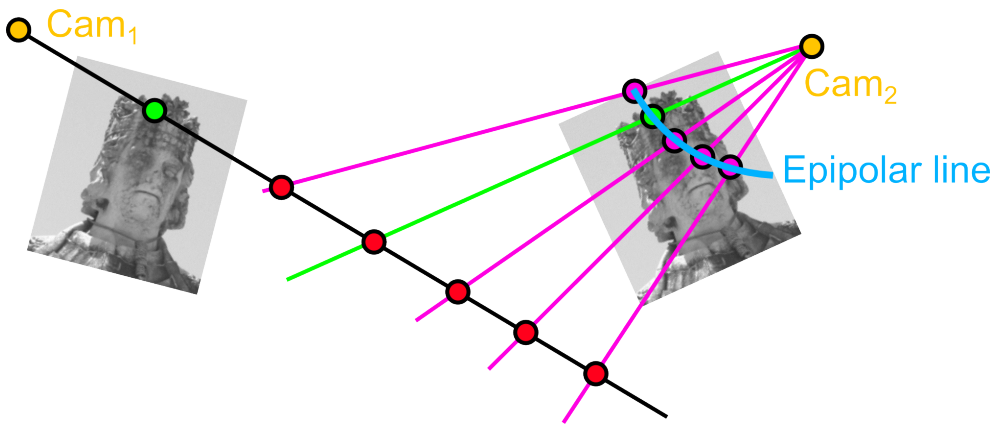

Figure 51: Different 3D solutions along the epipolar line.

For each pixel of the master image:

- we project an optical ray into "world coordinates".
- for different distances on the ray, we compute the position of the potential points.
- we project the points into the other images (the points on the images are on the **Epipolar Line**).
- for each point in each image we compute the correlation score.
- the distance yielding the best compounded score gives the position of the point.

3.9.3 Parallel method

For the plane method (see Fig. 53), we do not have a master image but rather an area of interest and a target planar resolution. The combination of these parameters creates a bi-dimensional grid of points for which the altitude needs to be determined.

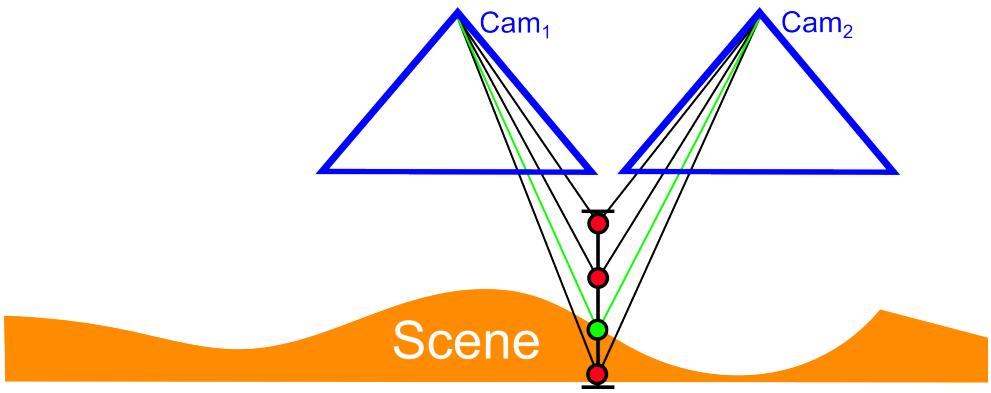

Figure 52: Projection in two images of the same planar point given four different altitudes.

For each point (x, y) of the grid:

- we project points (x, y, z) with $Alti_{min} < z < Alti_{max}$ into the images.
- we get templates from around each projected point, grouped by the z value that generated them.
- we score the correlation of all groups of templates (see Figure 53).
- the group yielding the best score gives the position of the point (see Figure 52).

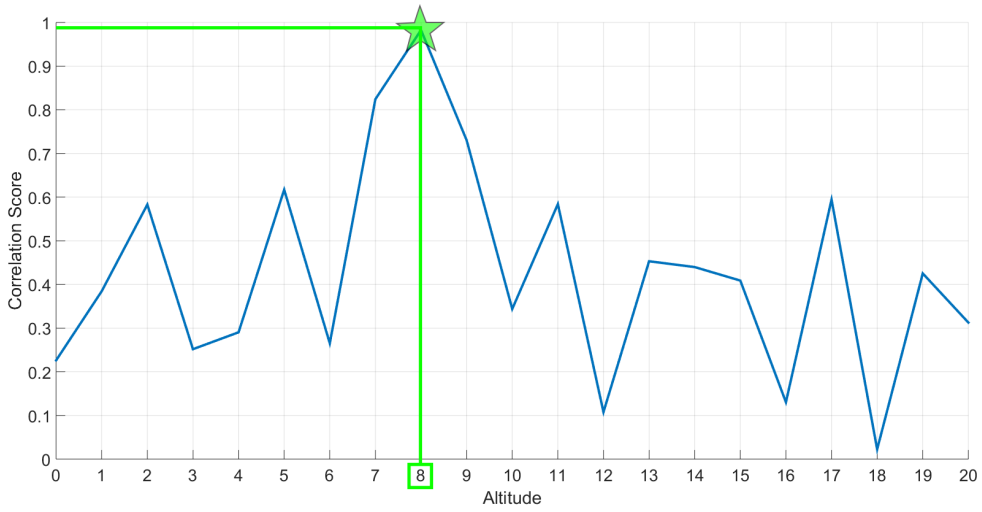

Figure 53: Mean correlation score of each group of templates for a given altitude. The best score determines the recorded altitude for that point; here, 8 m with a correlation score of 0.98.

3.9.4 Improving the speed and reliability of image correlation

3.9.4.1 Multi-resolution pyramidal correlation A popular method to increase both quality and speed of the processing is multi-resolution pyramidal correlation [Remondino et al., 2013]. The idea is to perform the correlation at a low resolution first, getting a rough model. Then the resolutions of both the model and the steps between candidate points on the projective ray/epipolar line are progressively refined, using the previous result as a guide to limit the search space (range of values tested along the epipolar line for example). Not going through the full search space at full resolution is primarily making the process faster, but also helps reduce noise, as potential mismatches are filtered out.

Figure 54: Scalled images (Top Left : full resolution ; Top Right : 1/2 ; Bottom Left : 1/4 ; Bottom Right : 1/8)

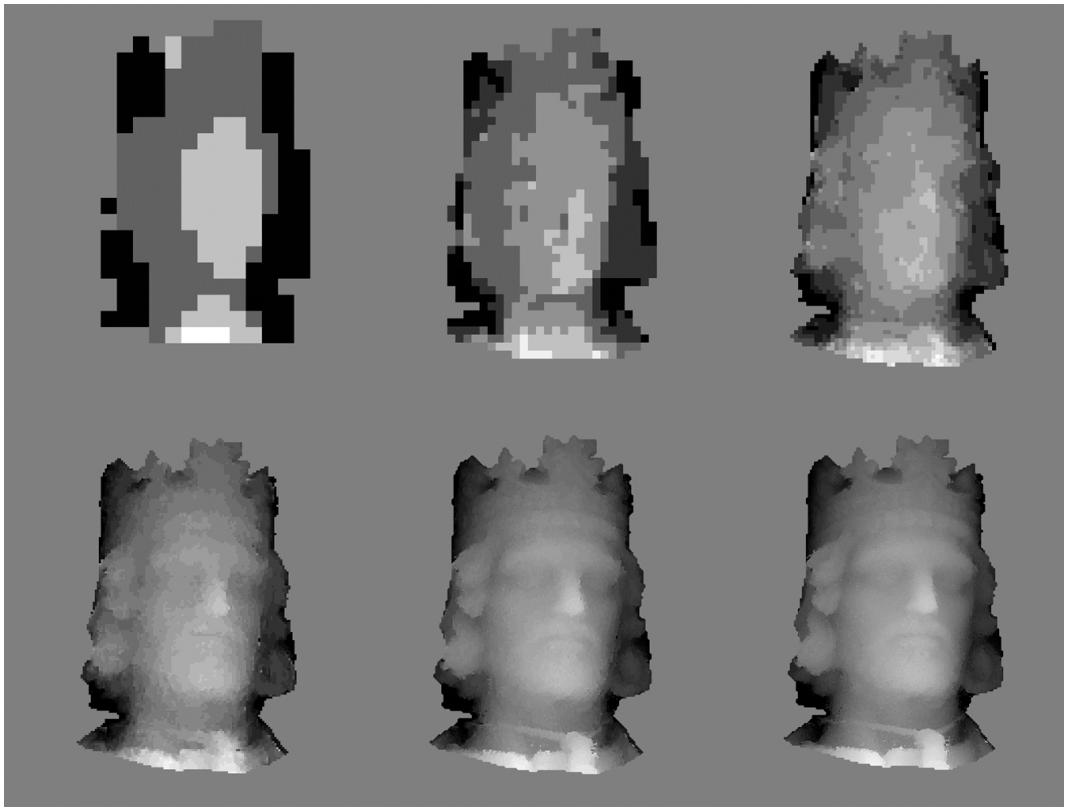

Figure 55: Products at different resolution (Top Left : 1/32 ; Top Middle : 1/16 ; Top Right : 1/8 ; Bottom Left : 1/4 ; Bottom Middle : 1/2 ; Bottom Right : full resolution

3.9.4.2 Multi-images matching While two images are enough to give geometric information, matching with more images gives and edge :

- 2 images : the result can only be trusted.
- 3 images : discordance can be identified and either the best match is kept or the results can be averaged.
- 4 images : outliers can be identified and excluded.
- having even more images gives a higher confidence in the results.

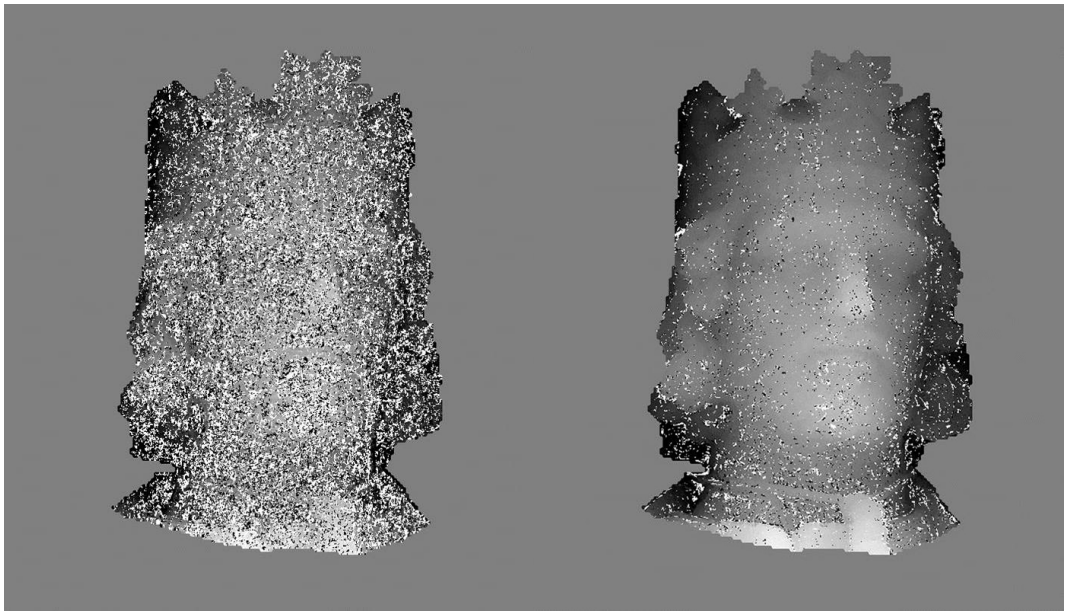

Figure 56: Left : 2 images ; Right : 4 images

3.9.4.3 Regularization In most cases, we have an a priori knowledge on the roughness of the surface (For instance : Are we going to see rolling hills or sharp cliffs/building?). We can therefore add a smoothness parameter to affect the correlation score. The weight of this parameter influences how much we "prefer" solution 1 to solution 2 in Figure 57.

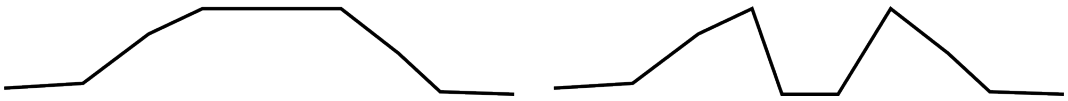

Figure 57: Left : solution 1, high smoothness a priori ; Right : solution 2, low smoothness a priori

We have the score altering formula :

$$Score_{smooth} = Score_{correlation} + \alpha/G$$

With :

- α the smoothness parameter ($0 < \alpha < 1$).
- G the resulting gradient in the model for the current potential solution (Normalized so $0 < G < 1$).

A strong alpha means that the surface should be smooth. The results of various alphas are shown in Figure 58.

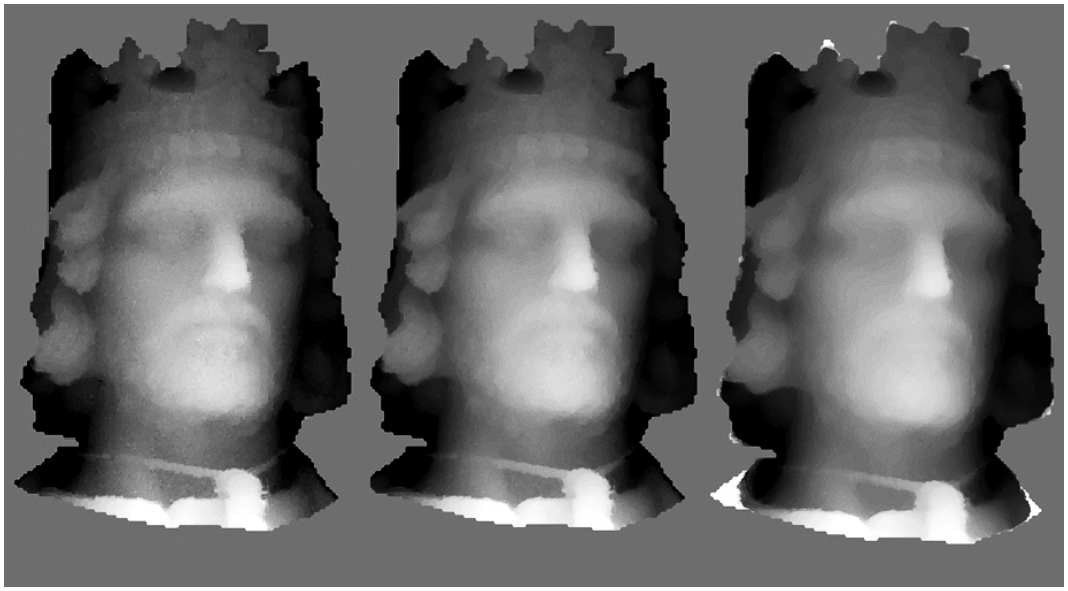

Figure 58: Results with varying alphas (0.01 ; 0.1 ; 0.5)

3.10 Orthorectification

Orthorectification is the process of correcting an image of its optical, terrain and pointing angle distortions. It is however easier to proceed backwards: for each point of the map, we want to have radiometric information. The orthorectification of an image can be done by applying the following algorithm for each point (X,Y) of the target grid (illustrated in Fig. 59):

1. Use a DEM to get the elevation value (Z) associated with (X,Y), using a geometric interpolation if the DEM grid is not the same as the target grid.
2. Project this 3D point (X,Y,Z) in the image to get image coordinates (i,j) through the function $\mathcal{O}$ (see Section 3.7).
3. Interpolate the radiometric value (RGB or Greyscale for instance) for the query point (i,j) in the image.
4. Record this value in the orthoimage (X,Y,Colour).

3.11 Mosaicing of orthoimages

Once individual orthoimages are computed, they need to be mosaicked into a single image of the whole area of interest. Since they are all in the same geometry, the mosaicking is quite straightforward as a first approximation: go through the cartographic space and average the values of the images available for all query points. An other common method is to use a Voronoï diagram [Voronoï, 1908; Okabe, 2016]: the image closest to the query point is chosen to give the colour.

However, both methods have limitations. For a number of reasons (most materials do not present a Lambertian reflectance so their bidirectional reflectance distribution function (BRDF) cannot be assumed to be constant), the same point in the terrain might not be the same colour in all images. This will create unwanted seams at the boundary of the areas covered by individual pictures. To come around these issues, the seam between two

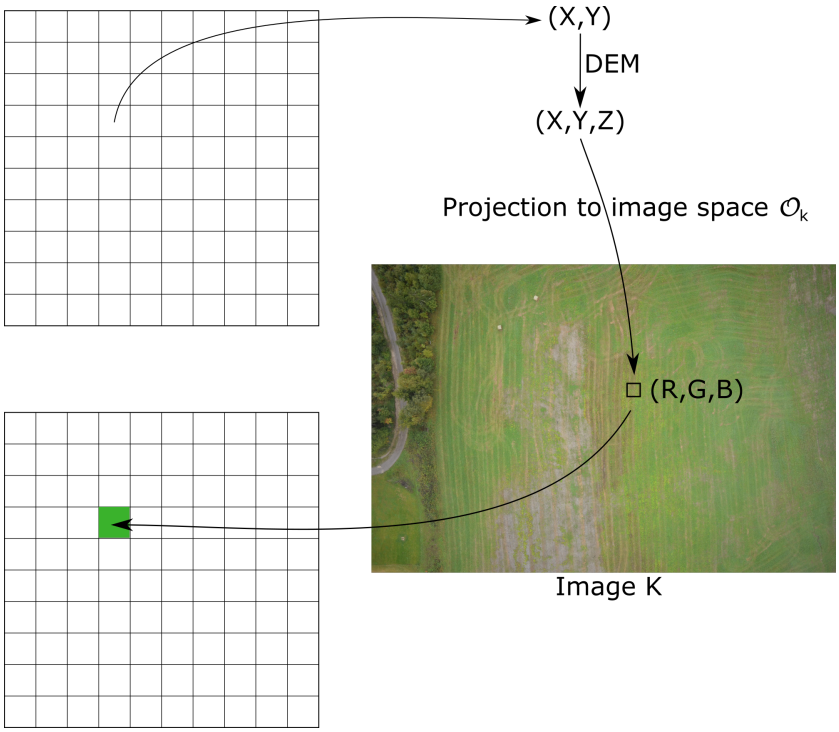

Figure 59: Schema for orthorectification. The image on the right is used to colour the grid map on the left.

pictures can be defined as following a natural boundary in the images (a line with strong contrast for instance). An other option is to fit corrections close to the seams to smooth the transitions.For instance, the mix between the two images could be 50% at the Voronoï diagram seem and evolve slowly to 100% of each image when getting away from the seem (see Figure 60, top right).

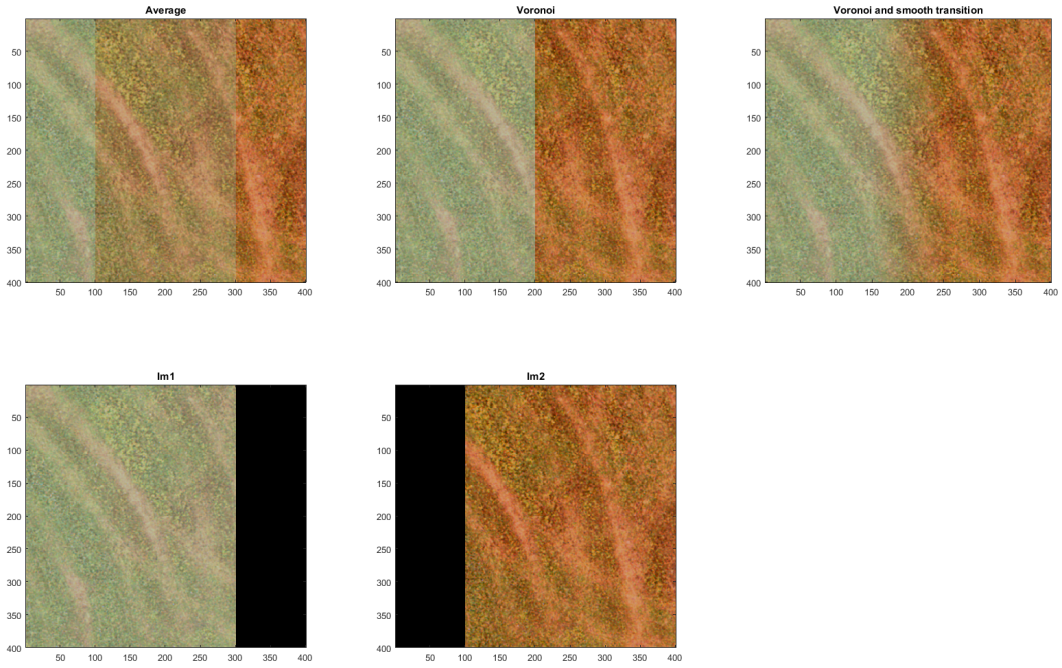

Figure 60: Different mosaicking approaches for two images (bottom row)

4 Videogrammetry

The field of videogrammetry has seen a development in the last few decades [Gruen, 1997] and is starting to be accessible [Rupnik and Jansa, 2014]. It is based on the principle of close range photogrammetry but uses the high frame rate of video from several cameras to reconstruct the position of points across the duration of the video.

Video can also be used in photogrammetry to increase the acquisition rate of a camera (the Panasonic GH5 can for instance capture 60 frames per second at a resolution of 3840*2160 pixels). Frames are then extracted from the video and create a very coherent flux of imagery for 3D modeling. Such a flux can be used for the creation of geometrically and temporally smoothed time-lapses called hyper-lapses [Kopf et al., 2014], or to reconstruct the 3D environment captured by the video. Movies not shot with this in mind can still be exploited in some cases, and the available video archived could be used to collect time-series of morphological data or archaeological information on lost buildings (similar to what was done by [Silver et al., 2016] on the cultural heritage of Syria), or simply for fun (see Fig. 61).

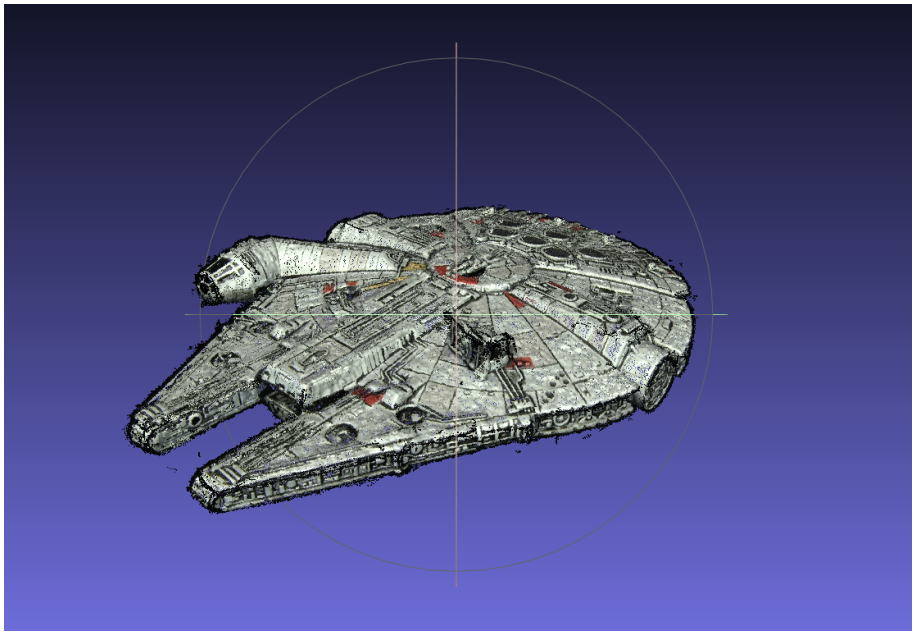

Figure 61: 3D model of the *Millennium Falcon* from the end scene of *Star Wars: Episode V - The Empire Strikes Back* (1980).

Bibliography

Agisoft LLC. Agisoft photoscan 1.3.2, August 2017. URL http://www.agisoft.com/.

Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359, 2008.

Fairchild, S. and Morton, E. Gyroscopic control of cameras and other optical devices, August 7 1928. URL https://www.google.com/patents/US1679354. US Patent 1,679,354.

Furukawa, Y. and Ponce, J. Accurate, dense, and robust multiview stereopsis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(8):1362–1376, 2010. doi: https://doi.org/10.1109/TPAMI.2009.161.

Girod, L. and Pierrot-Deseilligny, M. L'égalisation radiométrique de nuages de points 3d issus de corrélation dense. *Revue Française de Photogrammétrie et Télédétection*, 206: 3–14, 2014.

Gruen, A. Fundamentals of videogrammetry — a review. *Human Movement Science*, 16(2):155 – 187, 1997. ISSN 0167-9457. doi: http://dx.doi.org/10.1016/S0167-9457(96)00048-6. URL http://www.sciencedirect.com/science/article/pii/S0167945796000486. 3-D Analysis of Human Movement - II.

Gupta, R. and Hartley, R. I. Linear pushbroom cameras. *IEEE Transactions on pattern analysis and machine intelligence*, 19(9):963–975, 1997.

James, M. R. and Robson, S. Mitigating systematic error in topographic models derived from uav and ground-based image networks. *Earth Surface Processes and Landforms*, pages n/a–n/a, 2014. ISSN 1096-9837. doi: 10.1002/esp.3609. URL http://dx.doi.org/10.1002/esp.3609.

Kääb, A. *Remote sensing of mountain glaciers and permafrost creep*. Physische Geographie. Geographisishes Institut der Universität Zürich, 2005. ISBN 3 85543 244 9. URL https://books.google.no/books?id=Nx5NAQAAIAAJ.

Koenderink, J. J. and Van Doorn, A. J. Affine structure from motion. *JOSA A*, 8(2): 377–385, 1991.

Kopf, J., Cohen, M. F., and Szeliski, R. First-person hyper-lapse videos. *ACM Transactions on Graphics (TOG)*, 33(4):78, 2014.

Korona, J., Berthier, E., Bernard, M., Rémy, F., and Thouvenot, E. Spirit. spot 5 stereoscopic survey of polar ice: reference images and topographies during the fourth international polar year (2007–2009). *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(2):204–212, 2009.

Laussedat, A. *Mémoire sur l'emploi de la chambre claire dans les reconnaissances topographiques*. Mallet-Bachelier, 1854.

Lloyd, G. A. and Sasson, S. J. Electronic still camera, December 26 1978. US Patent 4,131,919.

Lowe, D. G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, November 2004. ISSN 0920-5691. doi: 10.1023/B:VISI.0000029664.99615.94. URL http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94.

McGlone, C., Mikhail, E., and Bethel, J. *Manual of photogrammetry, Fifth Edition*. American Society for Photogrammetry and Remote Sensing, 2004.

Meydenbauer, A. Die photogrammetrie. *Wochenblatt des Architektenvereins zu Berlin*, (49):471–472, Des 1867.

Morel, J.-M. and Yu, G. Asift: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, 2(2):438–469, 2009.

Niépce, J. N. Notice sur l'héliographie. *Historique et Description des Procédés du Daguerréotype et du Diorama*, pages 39–46, 1839.

Okabe, A. *Spatial Tessellations*. John Wiley & Sons, Ltd, 2016. ISBN 9781118786352. doi: 10.1002/9781118786352.wbieg0601. URL http://dx.doi.org/10.1002/9781118786352.wbieg0601.

Pierrot-Deseilligny, M., Jouin, D., Belvaux, J., Maillet, G., Girod, L., Rupnik, E., Muller, J., and Daakir, M. *MicMac, Apero, Pastis and Other Beverages in a Nutshell!* ENSG, IGN, Champs-Sur-Marne, France, August 2017.

Pix4D SA. Pix4d 3.2, August 2017. URL https://pix4d.com/.

Remondino, F., Spera, M. G., Nocerino, E., Menna, F., Nex, F., and Gonizzi-Barsanti, S. Dense image matching: Comparisons and analyses. In *2013 Digital Heritage International Congress (DigitalHeritage)*, volume 1, pages 47–54, Oct 2013. doi: 10.1109/DigitalHeritage.2013.6743712.

Rupnik, E. and Jansa, J. Off-the-shelf videogrammetry-a success story. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(3):99, 2014.

Rupnik, E., Daakir, M., and Pierrot Deseilligny, M. Micmac - a free, open-source solution for photogrammetry. *Open Geospatial Data, Software and Standards*, 2(14):1–9, 2017.

Saint-Amour, P. K. Applied modernism. *Theory, Culture & Society*, 28(7-8):241–269, 2011. doi: 10.1177/0263276411423938. URL http://dx.doi.org/10.1177/0263276411423938.

Silver, M., Rinaudo, F., Morezzi, E., Quenda, F., and Moretti, M. L. The cipa database for saving the heritage of syria. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 41, 2016.

Snavely, N. Bundler: Structure from motion for unordered image collections. *accessed*, 28(10):2010, 2010.

Snavely, N., Seitz, S. M., and Szeliski, R. Photo tourism: exploring photo collections in 3d. In *ACM transactions on graphics (TOG)*, volume 25, pages 835–846. ACM, 2006.

Tao, C. V. and Hu, Y. A comprehensive study of the rational function model for photogrammetric processing. *Photogrammetric engineering and remote sensing*, 67(12):1347–1358, 2001.

Voronoï, G. Nouvelles applications des parametres continus à la theorie des formes quadratiques. *Journal für die reine und angewandte Mathematik*, 134(3):27, 1908.