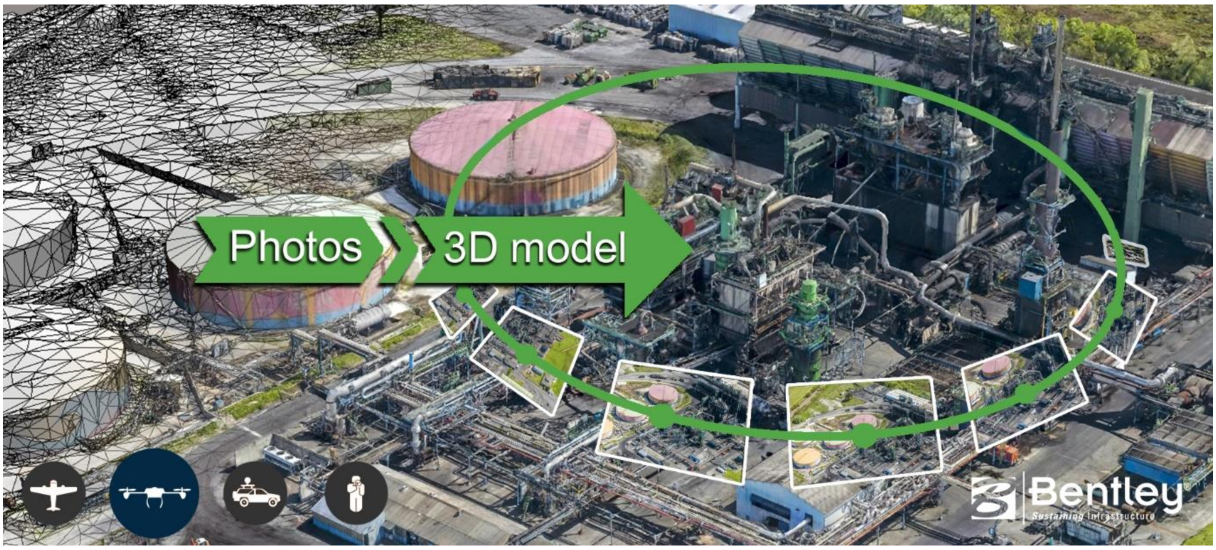
ContextCapture | Guide for photo acquisition



ContextCapture is automatically turning photos into 3D models, meaning that the quality of the input dataset has a deep impact on the output 3D model which is generated.

In this document, you can find useful information on how to take photos to obtain optimal results with ContextCapture.

Table of contents

How to setup your camera

Different types of camera can be used to capture an input dataset for ContextCapture. These can range from smartphones to the latest DSLR cameras and also include Camcorders. The quality of the resulting 3D models is dependent on the quality of the input photos and their spatial configuration.

To get the best results in ContextCapture, we recommend:

- A constant focal length during the acquisition: zoom "with your feet",
- Constant and homogeneous lighting.

You should avoid:

- Blurry photos: use adapted settings, and possibly a tripod under low lighting conditions.
- Flash light.
- Optical stabilization.

You must banish:

- Digital zoom,
- Any resizing/cropping/rotation of the input photos (turn off your camera's auto-rotation mode) ,

When Capturing Video the follow formats are recommended:

- Audio Video Interleave (AVI)
- MPEG-1/MPEG-2 (MPG)
- MPEG-4 (MP4)
- Windows Media Video (WMV)
- Quicktime (MOV)

For more information please refer to ContextCapture User Manual ("Preparing the imagery dataset").

Limitations

1. Enough photos must be taken to cover the entire object to successfully reconstruct. Every point of the scene must be captured in at least two adjacent photos.
2. Transparent or shiny parts cannot be properly reconstructed: glass, water, etc.
3. Parts of the object with a uniform aspect cannot be reconstructed: plain walls without texture, etc.
4. Photos looking at the same part must be taken from points of view that are neither not too similar nor too different.

How to capture objects

Capturing an object is among the simplest of tasks. Our recommendations for getting started with using ContextCapture are:

A little knowledge and some practice usage is required before it can be done properly. Note that that each part you want to reconstruct must be visible in at least two different photographs.

An easy way to fulfill this is to circle around the object taking small steps (ensure a minimum 60% overlap and a maximum angle difference of 15° is maintained between consecutive photos). Typically, you also want to get a uniform resolution. To achieve this, try to keep the camera at the same distance from the object during the acquisition process (Figure 1).

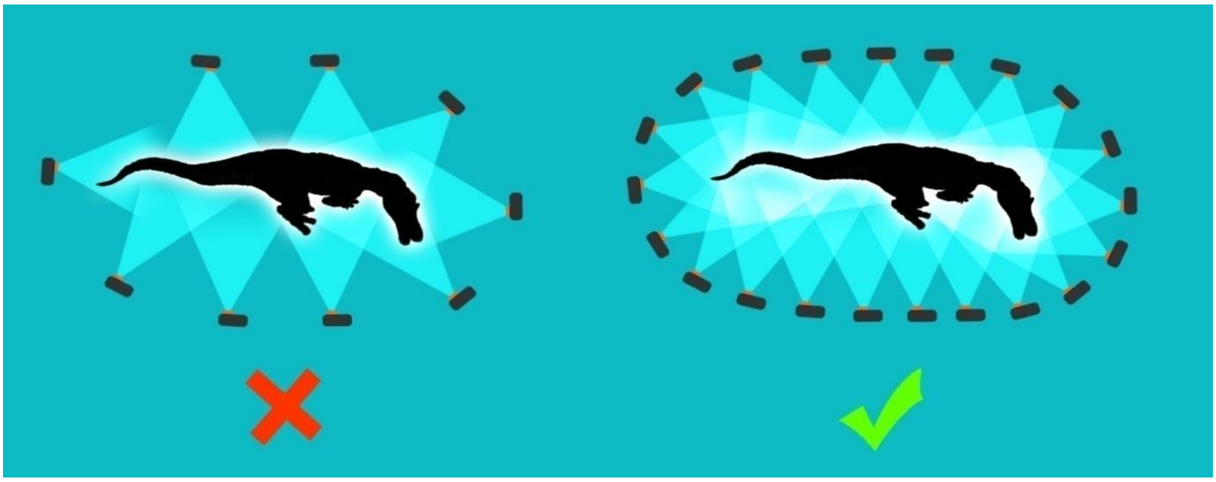


Figure 1: Object acquisition

If you want to obtain more detail in some parts of the 3D model, progressively take photos closer and closer to the object. Having photo resolutions that are too different may lead to a failure during the aerotriangulation. This is why it is recommended to have medium range photos which link the ones that are shot at close and long range. (Figure 2).

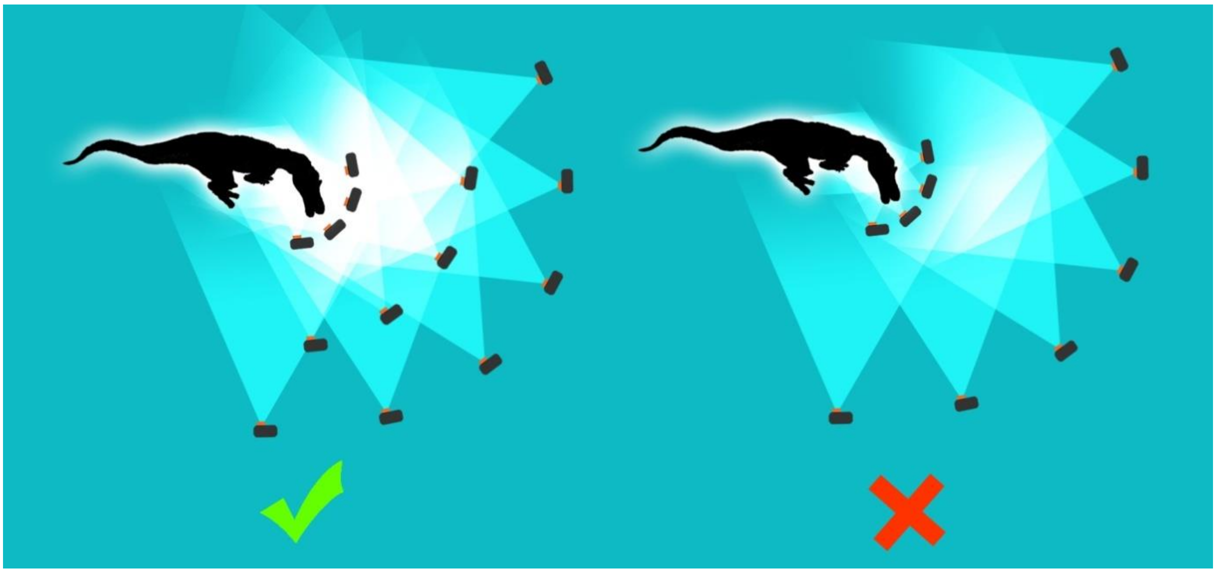


Figure 2: Capturing a part of the object in more detail

Ensure that all details of the object are scanned, by circling around it at different levels. This will reveal even the most hidden areas (Figure 3).

Figure 3: Circling at different levels

If the object has little texture, the aerotriangulation may fail due to an insufficient number of points of interest. To circumvent this, we recommend that the object be placed on a highly textured surface, for example a newspaper (Figure 5).

Figure 4: Using a highly textured surface

Capturing the bottom of the object is tricky because, in photogrammetry, the subject has to remain still throughout the acquisition. When turning the object upside-down to capture the bottom, the background will not be consistent with the rest of the dataset, which will result in a failure. There are several ways to deal with this difficulty:

- Taking photos of the object from all the needed points of view and masking the background manually in all the photos with a third-party image editing software (this process can be tedious).
- Creating a shooting "studio" / turntable method: this consists of a stand that is completely un-textured. For example, use a white stand over a uniform white background and turn the object instead of circling around it. You will then be able to capture any part of the object without being bothered by the background (Figure 4).

Figure 5: Turntable method

How to capture bodies and faces

Capturing a human face/body in order to create a 3D model can be really challenging. Indeed, the subject has to remain perfectly still during the whole acquisition, which is practically impossible to achieve with a living human.

For this reason professionals generally use RIGs (rigid camera systems). These systems are composed of several cameras synchronized to shoot pictures simultaneously. The acquisition of the full dataset is instantaneous, thus avoiding unwanted movements of the subject. The camera setup should follow the same principles as for object scanning. With the difference that increasing the number of pictures of your dataset means increasing the number of physical cameras, with an impact on the cost of the system.

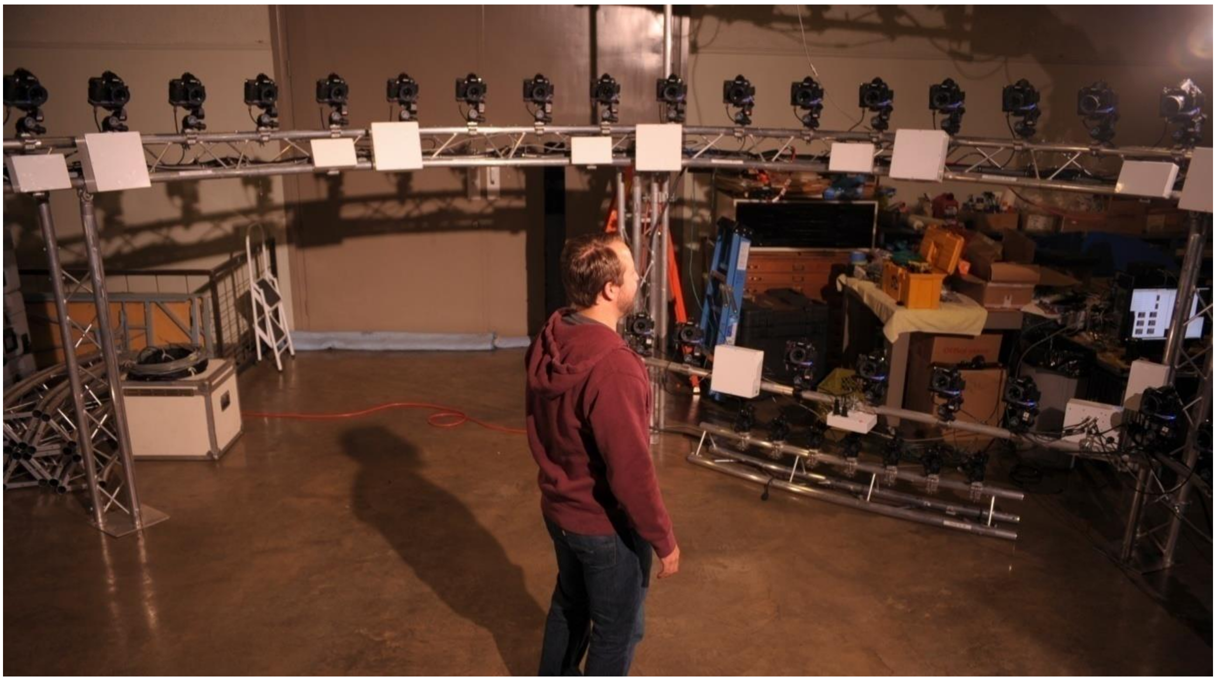


Figure 7: A rigid (abbr. RIG) camera system

If you cannot afford a rigid camera system, and need to capture human bodies and faces with a single camera, we recommend that you keep the acquisition time as short as possible, since the odds of movement by the subject increase over time. Hair is also quite complicated to model correctly, so be extra careful when shooting it and do not forget to take photos of the top of the head. You should also try to reduce the amount of background in your photos and try to get as much subject as possible (Figure 6). The same applies to any type of scene.

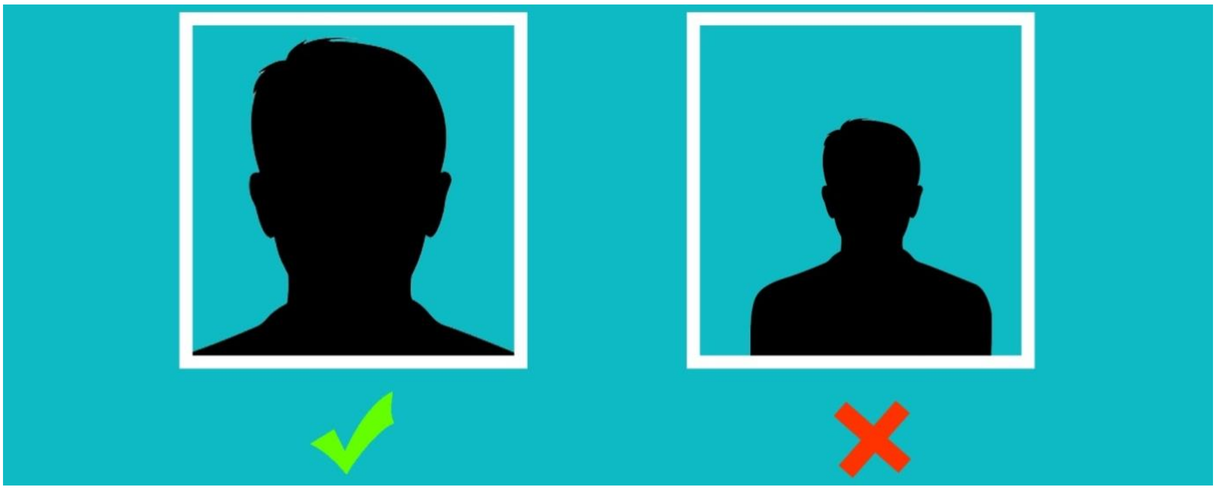


Figure 6: Avoid large background areas

How to capture façades and buildings

Some rules have to be followed in order to ensure a good acquisition:

- Follow a specific shooting order to avoid missing parts.
- Shoot the same part from different points of view at least 3 times.
- Limit the angle between two consecutive photos to a maximum of 15°.

The simplest way to capture a facade is to shoot photos from different evenly spaced stations, with different angles (Figure 8). Ensure a 60% overlap between two consecutive front views and a maximum angle of 15° for the side views.

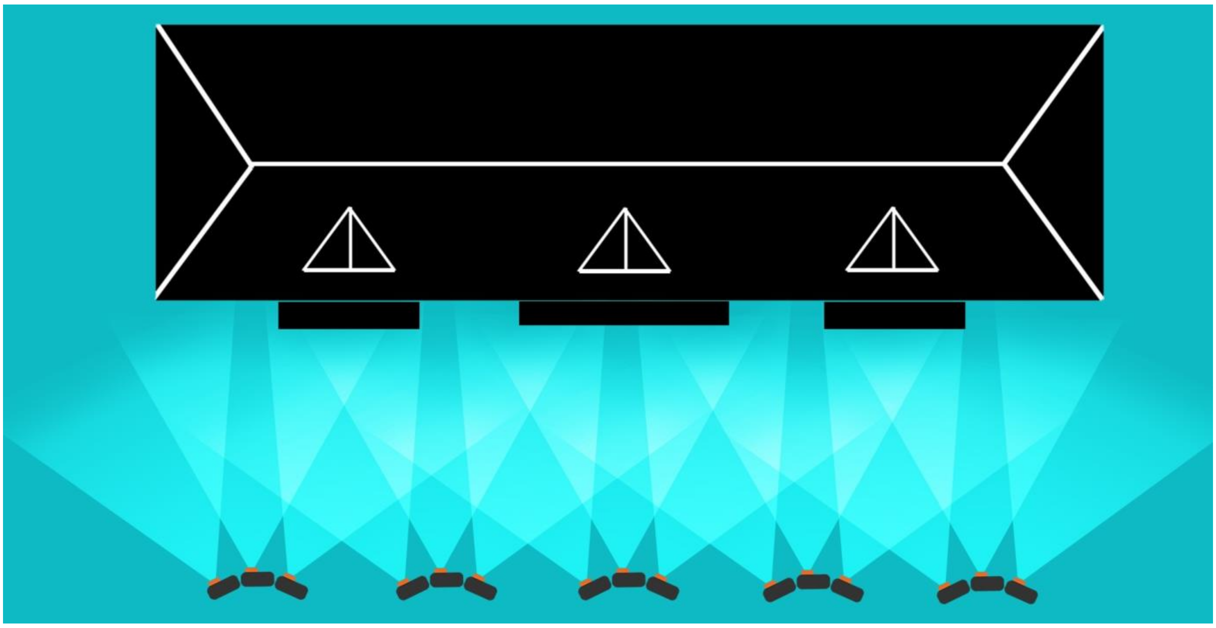


Figure 8: Facade scanning

If you want to reconstruct the complete building and not only a single facade, make sure to respect the maximum angle of 15° between two consecutive pictures when going around corners (Figure 9). This way, you will help ContextCapture to connect the two faces of the building.

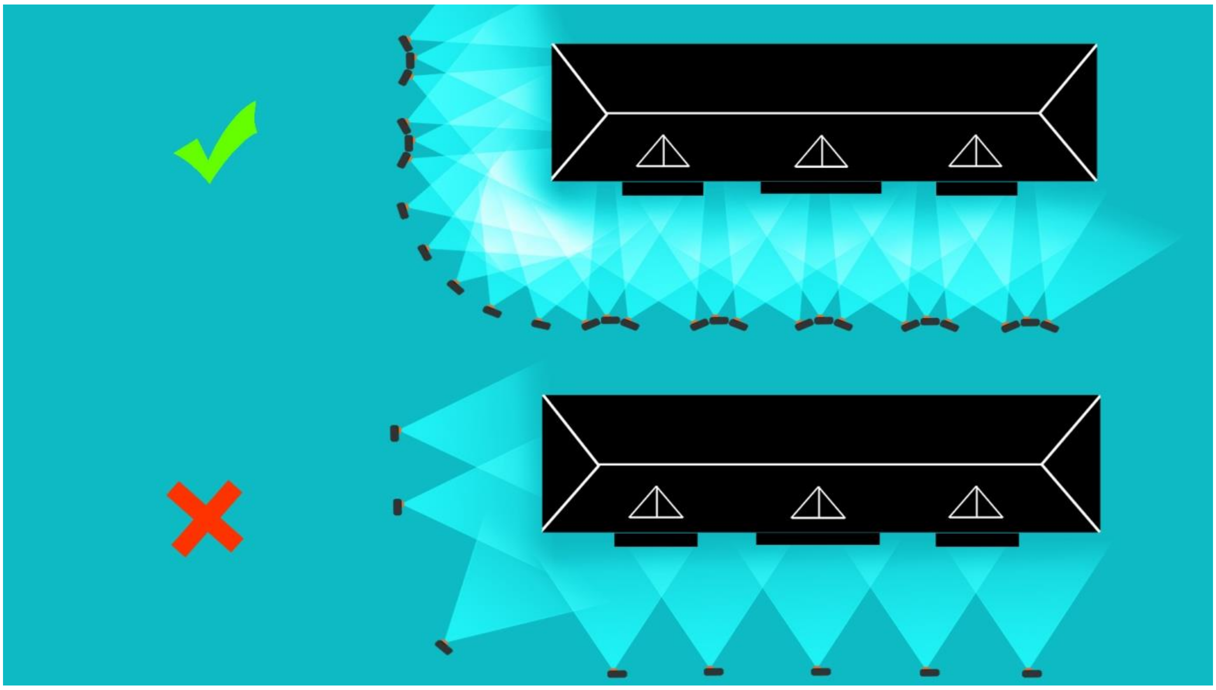


Figure 9: Turning around a building corner

If the full height of the building cannot be photographed due to a lack of distance and/or because the building is too tall, you will have to reproduce the pattern of Figure 9 at several heights. This requires a lift, a mast or a UAS/UAV/drone. Otherwise, just take successive photos from bottom to top, respecting the 60% overlap rule. In such case, please note that ContextCapture may lack information to properly reconstruct the high parts and the roof.

When using a UAS, we recommend to circle around the building at different levels to limit hidden areas (Figure 10). In this case, you can complete your aerial dataset with ground photos.

Remember that reflective, shiny or transparent parts may be hard or even impossible to reconstruct.

Figure 10: Circling around a building at different heights

How to capture interiors

Reconstructing interiors using photogrammetry is a difficult task. The short distance from the subject and the numerous objects creating masks drastically increase the number of photos needed to reconstruct the scene properly.

Figure 11 shows the correct acquisition pattern to capture interior scenes. To get the maximum distance from the scene, stand close to the wall and shoot the opposite side of the room. A common mistake is to stand in the center of the room and shoot in a panoramic fashion.

You may also need to reproduce the pattern shown in figure 11 at several heights. If tables or other pieces of furniture stand in the room, you will need extra photos to capture the bottom part of it.

Another common issue is the lack of texture on the walls. This may lead to holes in the 3D model, or even a failure during aerotriangulation (see "Limitations" p.3).

Use of a Fish-eye lense could be useful in circumstances where the distance from the scene (such as interiors) is limited.

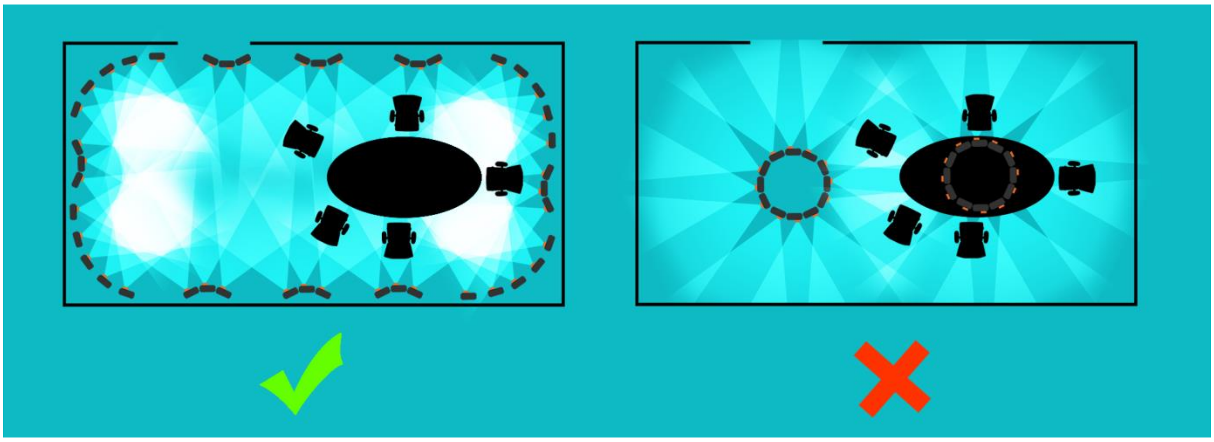


Figure 11 : advised interior acquisition

UAS/UAV/drone acquisition setups

Aerial dataset acquisition from UAS/UAV/Drones (Unmanned Aerial System/Vehicle) can be used for different purposes including ortho-photography, DSM production, inspection, single building modeling or neighborhood modeling (city-scale modeling requires acquisition systems with more autonomy).

The standard acquisition setup differs from one application to another and it should be adapted to obtain the best results for each one.

When precision really matters, we recommend the use of quality lenses with small distortion and preferably the choice of a camera with a fixed focal and a large sensor size. For a given number of megapixels per photo, a camera with a larger sensor size will produce better quality photos. This being said, the choice of the camera is often ruled by the maximum payload of your UAV.

1. 3D model precision

While professional UAS/UAV/Drones often come with dedicated flight planning software (automatically creating the best flight plan for a given camera, lens and desired ground sample distance), it is interesting to know how to calculate the expected precision of your results.

Topographic survey specifications always include the precision of the results. The following formula will help you estimate the expected precision of the output from ContextCapture.

- **Ls** is the greater size of the sensor (mm)
- **D** is the distance between the camera and the subject (m).
- **f** is the focal length of the camera (mm)
- **L** is the greater size of the photograph (Px)
- **R** is the spatial ground resolution of the photos in (m/Px)
- **P** the precision of spatial positioning of the vertices of the 3D mesh.

$$R = \frac{Ls \times D}{f \times L}$$

$$P = 3 \times R$$

Because **Ls**, **f** and **L** are usually fixed for a given camera, the only way to improve accuracy is to decrease the flight height. This will also result in taking more photos to cover the same area.

P is the relative precision of the 3D mesh. The absolute precision of your model only make sense if it is geo-referenced.

In ContextCapture, your model can be geo-referenced either by using the imported 3D positions of the photos (through EXIF tags or import file), or by adding control points (most of the spatial reference systems are supported).

The absolute precision will then depend on the relative precision of the 3D mesh and on the precision of the control points (or photo 3D positions).

2. Ortho-photographs and DSM production

Ortho-photo and DSM production does not necessarily require the acquisition of oblique views. Nadir photos with enough overlap are sufficient to process it (many UAS/UAV/drone acquisition systems are not able to capture oblique imagery).

In such cases, we recommend 80% of forward overlap and 60% of side overlap for a flight-line acquisition (Figure 12).

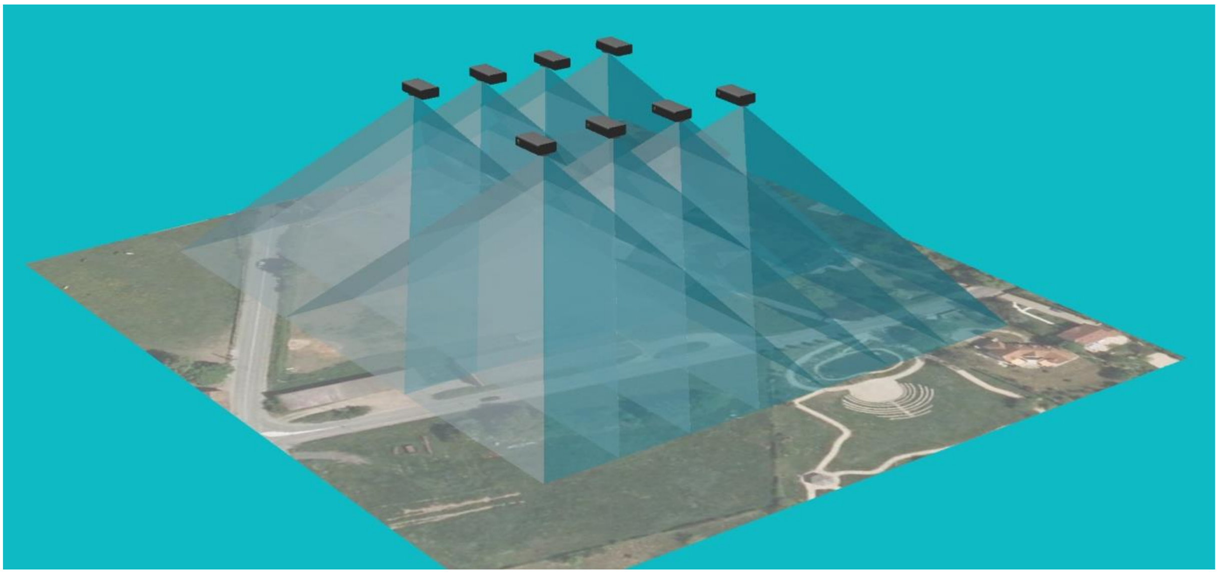


Figure 12 : overlap for Ortho/DSM

A survey of ground control points, evenly spread around the area of interest is also recommended.

<u>Known issue during the self-calibration of radial distortion:</u>

For acquisition setups consisting of nadir photos only, you may encounter a curve effect after the aerotriangulation step (Figure 13). Acquisitions consisting in straight lines along a single axis over a fairly flat surface are likely to be critical.

This curve effect is due to a critical configuration that leads to ambiguities during the computation of the radial distortion parameters. This flaw is shared by every structured from motion (SfM) software. This is analyzed in the following research paper: Changchang Wu, "Critical Configurations for Radial Distortion Self-calibration", CVPR 2014.

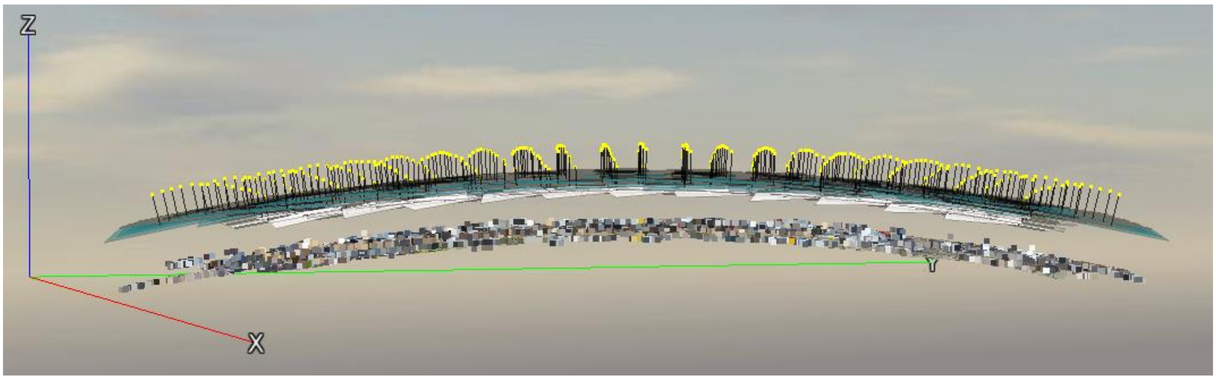


Figure 13 : curve effect due to ambiguities during radial distortion self-calibration

<u>How to fix it/prevent it from happening:</u>

The best way to avoid it, is to calibrate your camera before the flight.

In the latest release of ContextCapture, we are now able to store a camera calibration in a "user camera database" (right click on the photogroup -> "add camera model to database" / "get camera model from database").

Please note that the parameters of your camera are subject to change over time (especially for non-metric cameras). Therefore it is recommended that you calibrate your camera as often as possible, and preferably, right before the flight.

The workflow for this solution is the following:

- Before the flight, shoot about 20-30 photos around a highly textured and geometrically complex object,

with the same camera properties that will be used during the flight.

- Perform your aerial acquisition.
- Perform the AT of the 20-30 photos of your object (non-critical setup) and add the resulting camera calibration to the database (Right click on the photogroup -> “add camera model to the database”, Figure 14).
- You can now re-use this camera calibration in your aerial project: right click on the photogroup -> “get camera model from the database” to import your adjusted camera model and submit the AT. Do not forget to prevent the software from adjusting those parameters again (Figure 15).

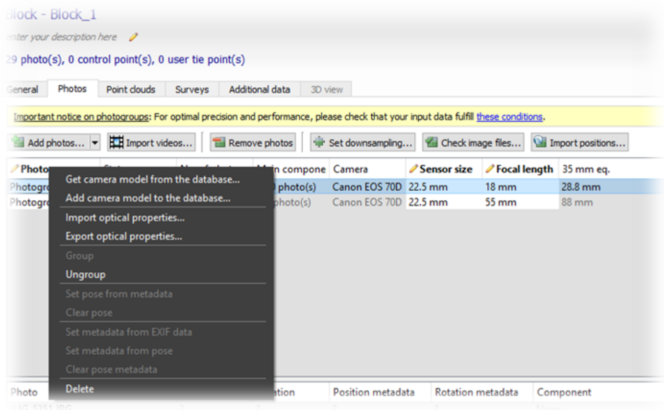


Figure 14: add camera model to database” / “get camera model from database (right-click on the photogroup)

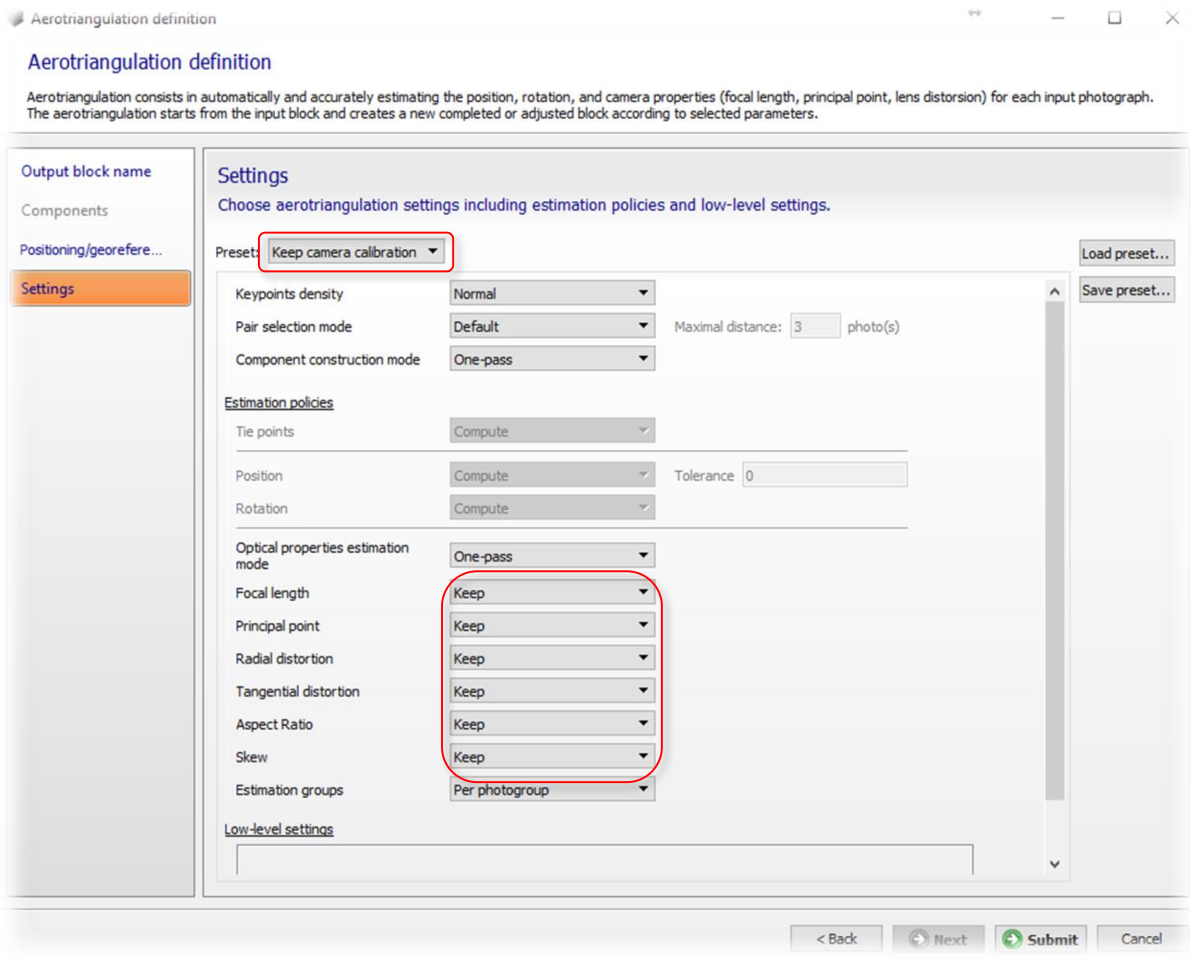


Figure 15: Keeping the imported optical properties during the AT step

The second option is to add some additional views to your aerial, with more angles, along different axis or with different flight heights. This solution is less recommended.

3. Full 3D modeling

Getting complete 3D models with a UAS/UAV/Drone can be quite complex and requires a large amount of photos including nadir and oblique. Because most UAS/UAV/Drone are only able to load one camera at a time, several flights are needed with different camera orientations to gather all the points of view. Please have a look at the next section for further information ("Aerial multi-camera system dataset acquisition").

Aerial multi-camera system dataset acquisition

Generating a full 3D model of large areas such as cities requires an exhaustive dataset acquisition. This usually requires a multi-camera system that provides nadir and oblique photos in different directions, thus limiting the number of flights needed to cover the whole area.

This type of multi-camera systems may be home-made or bought from a company commercializing its own acquisition systems (VisionMap, IGI, Microsoft Vexcel, Icaros, Visual Intelligence, TrackAir, etc.)

We recommend the following parameters:

- 1 Nadir and 4+ oblique cameras setup.
- 80% forward overlap and a 60% side overlap.
- Having the approximate same ground spatial resolution from nadir and oblique photos.
- Oblique angles lower than 45° from vertical.

If your scene includes high buildings and narrow streets, you should choose a slightly lower oblique angle (around 30° from vertical). By doing so, you will better capture the streets and the bottom of the buildings, avoiding holes in the final model.

Figure 16 : dataset acquisition over Marseille, France

The distance between flight lines and the shooting frequency will be determined by the height of flight to ensure the required overlaps between photos (for a given camera).

Example

For this example we will consider an UltraCam Osprey camera system from Vexcel. We are going to learn how to calculate the flight height, how to calculate the distance between flight lines, the shooting rate and the overlap and ground resolution of the oblique photos.

- 1 Nadir camera
  - Sensor size: 70.044*45.084 mm
  - Focal length: 80 mm
  - Photo dimensions: 11674*7514 Pix
  - Boresight angle: vertical
  - Orientation: landscape (according to flight direction).

- 2 oblique cameras oriented forward and backward
  - Sensor size: 53.4*39.9 mm
  - Focal length: 120 mm
  - Photo dimensions: 8900*6650 Pix
  - Boresight angle: α = 45° from vertical
  - Orientation: landscape

- 2 oblique cameras oriented left and right
  - Sensor size: 53.4*39.9 mm
  - Focal length: 120 mm
  - Photo dimensions: 8900*6650 Pix
  - Boresight angle: α = 45° from vertical
  - Orientation: portrait

Let's consider the following values (Figure 17):

Ls is the greatest size of the sensor (mm)
ls is the smallest size of the sensor (mm)
D is the distance between the camera and the subject (m).
H is the flight height
m is the length of the greatest side of rectangle covered by the photo on the ground
n is the length of the smallest side of the rectangle covered by the photo on the ground
f is the focal length of the camera (mm)
L is the greatest size of the photograph (Px)
l is the smallest size of the photograph (Px)
R is the spatial resolution of the photos in (m/Px)
P the precision of spatial positioning of the vertices of the 3D mesh.

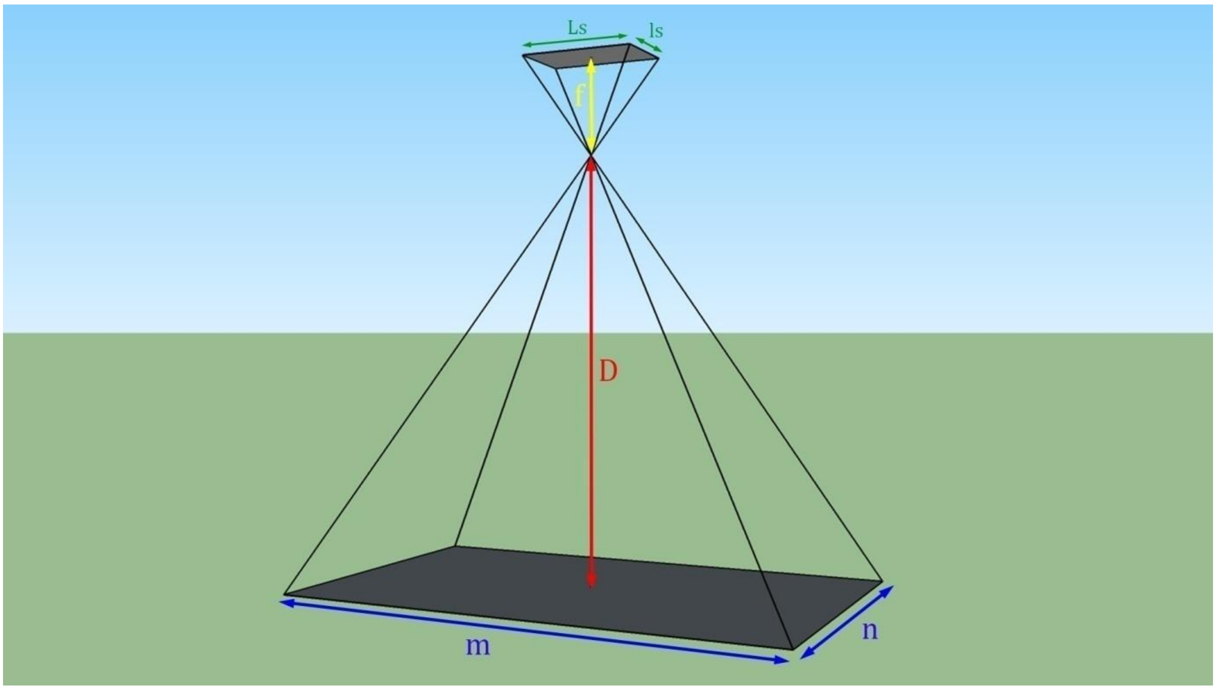


Figure 27 : photo projection on the ground

1. Estimating the flight height

The flight height will depend on the wanted precision of the model. For this example we will consider a wanted precision of 15cm on an entire city area. The calculation will be based on the Nadir camera, and we will double check with the oblique cameras later (multi-camera system manufacturers usually build their systems so the ground spatial resolution is about the same between nadir and oblique cameras).

First, calculate the ground spatial resolution required.

$$\mathrm{P} = 3 \times \mathrm{R}$$

$$R = \frac{P}{3}$$

$$\mathrm{R} = 0.05\ \mathrm{m/Px}$$

Now, we will calculate the surface on the ground that needs to be covered by the photos to achieve this resolution (Figure 17).

$$\mathrm{m} = \mathrm{R} \times \mathrm{L}$$

$$\mathrm{m} = 0.05 \times 11674 = 584\ \mathrm{m}$$

$$\mathrm{n} = \mathrm{R} \times \mathrm{l}$$

$$\mathrm{n} = 0.05 \times 7514 = 376\ \mathrm{m}$$

We can now calculate the corresponding flight height (for Nadir photos, the flight height H is equivalent to the distance from subject D)

(Because f is not significant over D, some simplifications are done in the formulas provided in the rest of the document)

$$H = D = \frac{f \times R \times L}{Ls} = f \times \frac{m}{Ls}$$

$$H = D = 80 \times \frac{584}{70} = 670 \text{ m}$$

2. Estimating the flight lines setup and shooting frequency

The flight line setup depends on the required overlap. We will consider a longitudinal overlap of 80% and a side overlap of 60% for the Nadir photos in this example (Figure 18).

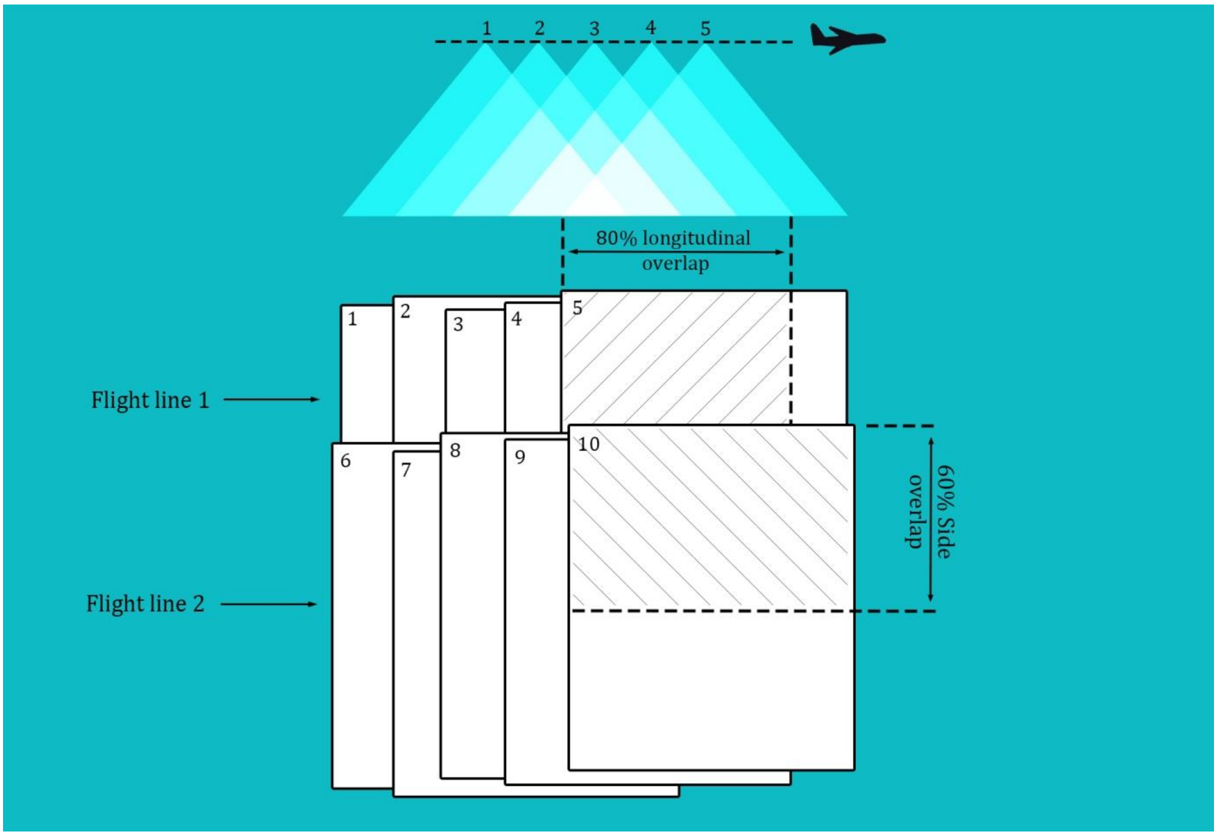


Figure 38 : longitudinal and side overlap

The distance between two consecutive photos in the same flight line is given by:

$$\Delta = 376 \times \frac{20}{100}$$

$$\Delta = 75 \text{ m}$$

The distance between two flight lines is given by:

$$\Delta = 584 \times \frac{40}{100}$$

$$\Delta = 233 \text{ m}$$

3. Checking the oblique setup

- **Ground spatial resolution**: we will consider the mean ground sample distance (because of the large angle from vertical, the ground spatial resolution can be quite different from one side of the photo to another).

We will consider the following setup: forward and backward oblique in landscape orientation (Figure 19).

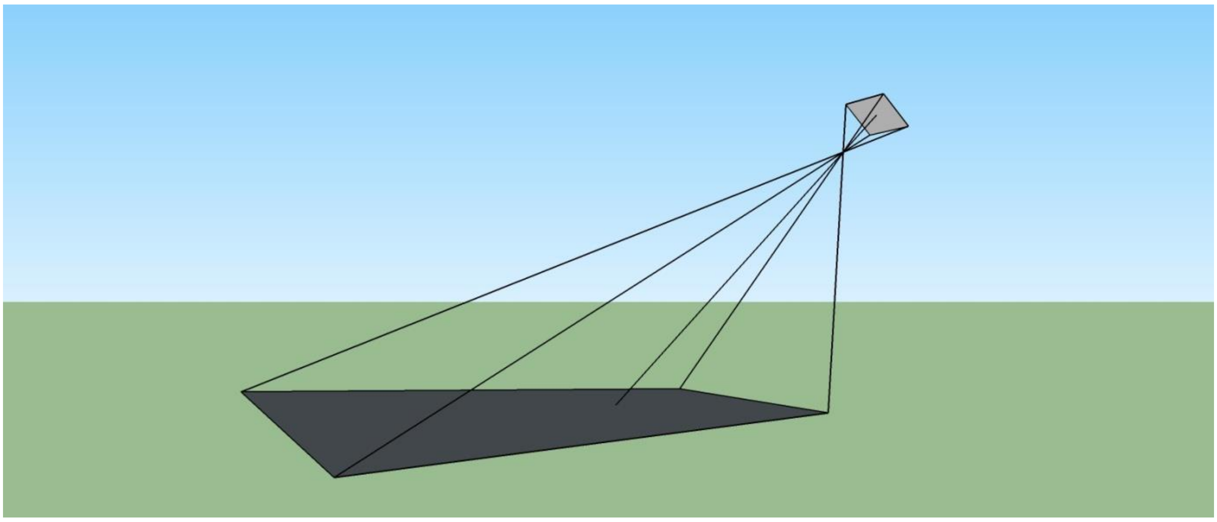


Figure 49 : Forward Oblique in landscape orientation

We will first calculate the dimension **n** of the projected photo on the ground (Figure 20).

First, we need angle **a**:

$$a = \tan^{-1}\left(\frac{ls}{2 \times f}\right)$$

$$a = 9.4°$$

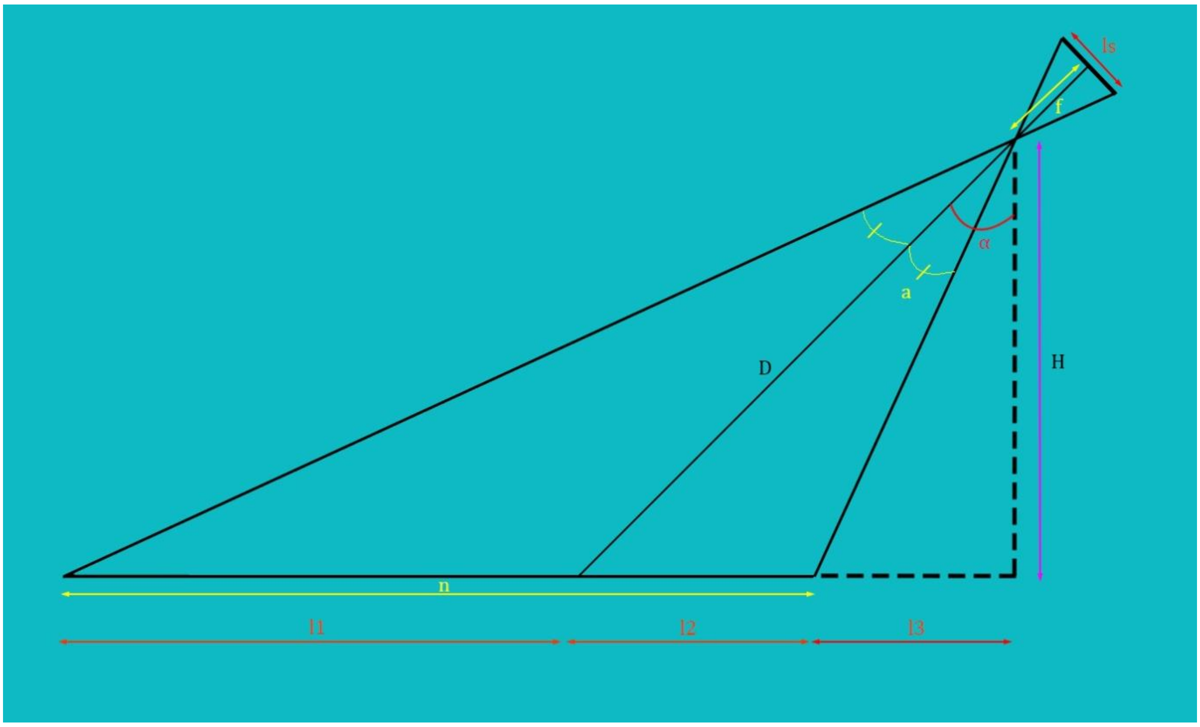


Figure 20: geometry

We can now calculate l1, l2, l3 and n:

$$l3 = H \times \tan(\alpha - a)$$
$$l3 = 480 \text{ m}$$

$$l2 = H \times \tan(\alpha) - l3$$
$$l2 = 190 \text{ m}$$

$$l1 = H \times \tan(\alpha + a) - l3 - l2$$
$$l1 = 266 \text{ m}$$

$$n = l1 + l2 = 456 \text{ m}$$

The mean ground spatial resolution is given by:

$$R = 456/6650 = 0.07 \text{ m/Pix}$$

The mean ground spatial resolution for the bottom half of the photo is given by:

$$R = l2/(l/2) = 190/3325 = 0.057 \text{ m/Pix}$$

The mean ground spatial resolution for the upper half of the photo is given by:

$$R = l1/(l/2) = 266/3325 = 0.08 \text{ m/Pix}$$

The results show that the ground spatial resolution is quite consistent for all the cameras of the system.

- Longitudinal Overlap : we can now check the longitudinal overlap between two consecutive photos from the same oblique camera (forward or backward) :

$$\text{LongitudinalOverlap} = \left(\frac{456 - 75}{456}\right) \times 100 = 83\%$$

- Side Overlap : we can now check the side overlap from two different flight lines between two photos from the same oblique camera (forward or backward) :

We will first now calculate D and m (Figure 18 and Figure 15):

$$\text{D} = \frac{\text{H}}{\cos(\alpha)} = 948 \text{ m}$$

$$\text{m} = \frac{\text{Ls} \times \text{D}}{\text{f}}$$

$$\text{m} = 421\text{m}$$

$$\text{SideOverlap} = \left(\frac{421 - 233}{421}\right) \times 100 = 44\%$$

We can notice that the side overlap is a bit low. You can increase it by reducing the distance between the flight lines. Anyway, ContextCapture is able to find matches between the oblique and nadir photos, thus lowering the effect of this small side overlap between the oblique photos.

The same calculation can be applied to estimate the overlap for the other oblique cameras.

CONCLUSION

Depending on the subject, the wanted precision, the purpose of the 3D modeling, and the acquisition device, the survey can be totally different from one project to another. Each scenario has to be carefully studied before going on the field and starting the acquisition. Else you will likely miss important photos and end up with incomplete or bad results.

This document intended to show you the best practices for different applications, using several devices including hand-held, UAVs-mounted or airborne cameras. Remember that the photo is at the basis of all the computation. The better the quality and the spatial sampling of your photos are, the better the 3D reconstruction will be.