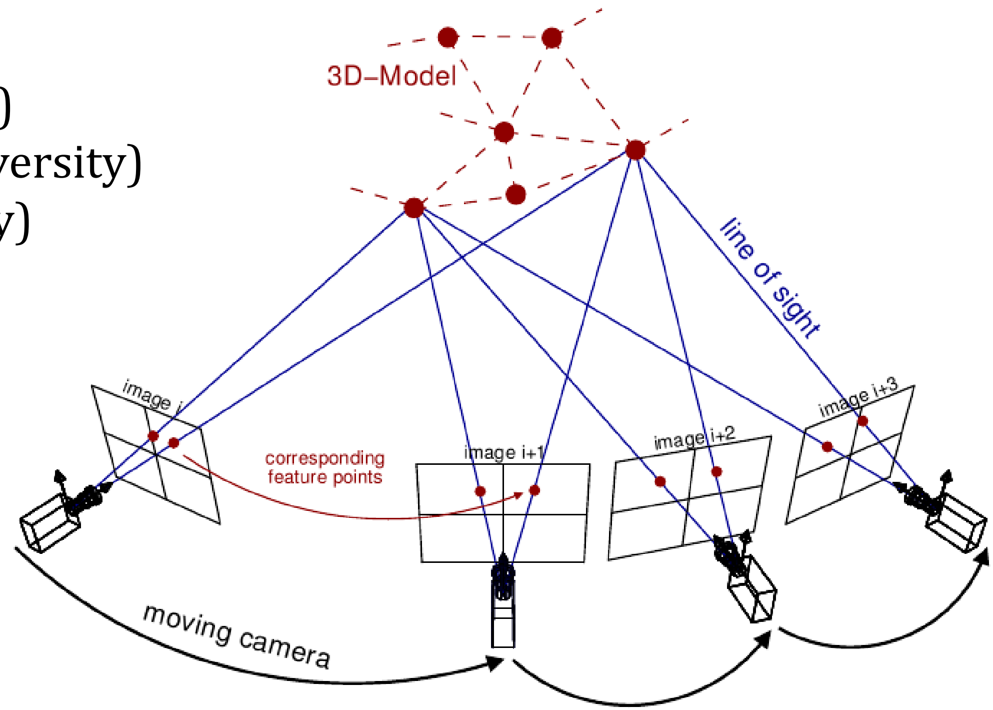
Introduction to Structure-from-Motion

Edwin Nissen (Colorado School of Mines)
J Ramon Arrowsmith (Arizona State University)
Chris Crosby (UNAVCO/OpenTopography)



- **What is Structure-from-Motion?**
- **High resolution topography: scientific motivations**

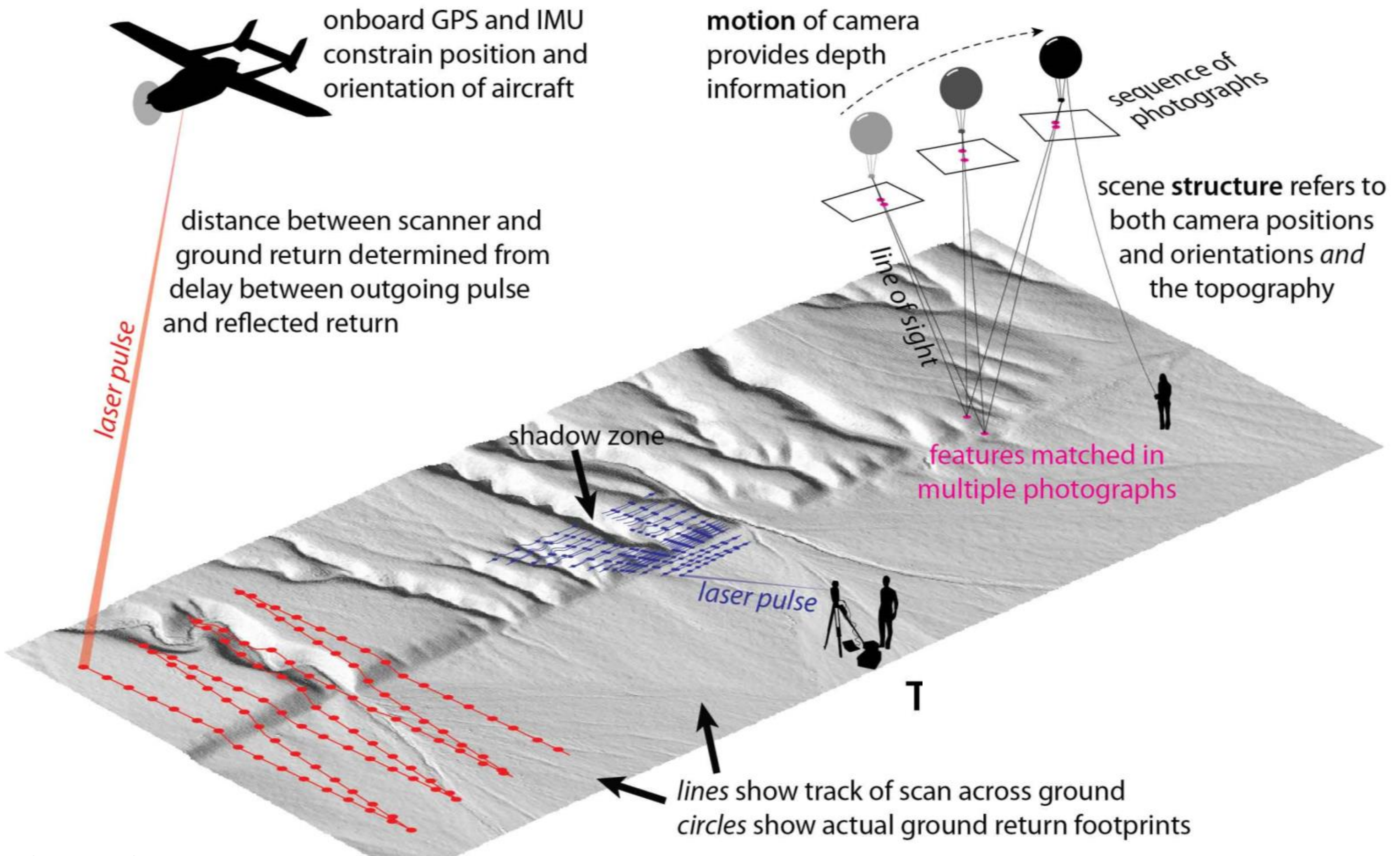
~500 points/m^2 coloured point cloud along a ~1 km section of the 2010 El Mayor-Cucapah earthquake rupture generated from ~500 photographs captured in 2 hours from a helium blimp

Lidar (ALS, TLS, MLS)

- Expensive laser equipment required
- Works in densely-vegetated landscapes
- Uses precise time-of-flight measurements but prone to artifacts from GPS and IMU

Structure-from-Motion

- Requires only a cheap camera
- Coloured points & orthophoto for texture mapping
- Back-solves for camera parameters; warping artifacts are a common problem but easily mitigated



Johnson *et al.* (2014), *Geosphere*

Where it all started...

The original idea

Proc. R. Soc. Lond. B. **203**, 405–426 (1979)
Printed in Great Britain

The interpretation of structure from motion

By S. Ullman
Artificial Intelligence Laboratory, Massachusetts Institute of Technology,
545 *Technology Square* (*Room* 808), *Cambridge, Massachusetts* 02139 *U.S.A.*

(*Communicated by S. Brenner, F.R.S. – Received* 20 *April* 1978)

The interpretation of structure from motion is examined from a computional point of view. The question addressed is how the three dimensional structure and motion of objects can be inferred from the two dimensional transformations of their projected images when no three dimensional information is conveyed by the individual projections.

Where it all started…

The algorithm that powers SfM

Proc. of the International Conference on Computer Vision, Corfu (Sept. 1999)

Object Recognition from Local Scale-Invariant Features

David G. Lowe

Computer Science Department
University of British Columbia
Vancouver, B.C., V6T 1Z4, Canada
lowe@cs.ubc.ca

Abstract

An object recognition system has been developed that uses a new class of local image features. The features are invariant to image scaling, translation, and rotation, and partially invariant to illumination changes and affine or 3D projection.

Where it all started…

• The **Scale Invariant Feature Transform (SIFT)** (Lowe, 1999) allows corresponding features to be matched even with large variations in scale and viewpoint and under conditions of partial occlusion and changing illumination

Where it all started…

First use of the SIFT algorithm to generate large point clouds

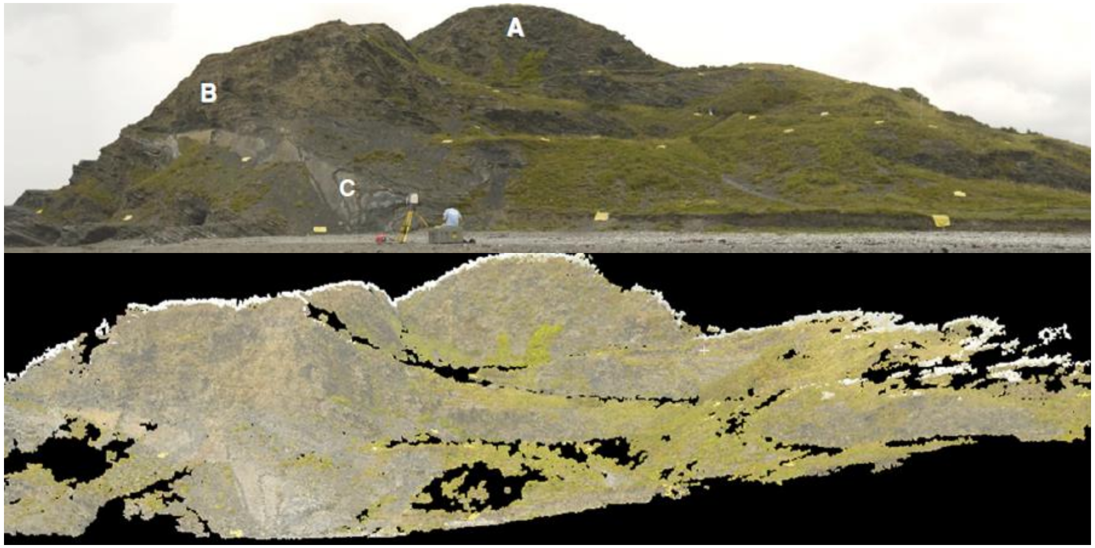
Snavely *et al.* (2006). Photo Tourism: Exploring Photo Collections in 3D, *ACM Transactions on Graphics*

Snavely *et al.* (2007). Modeling the World from Internet Photo Collections, *International Journal of Computer Vision*

Using photographs from a **moving** camera (or cameras)…

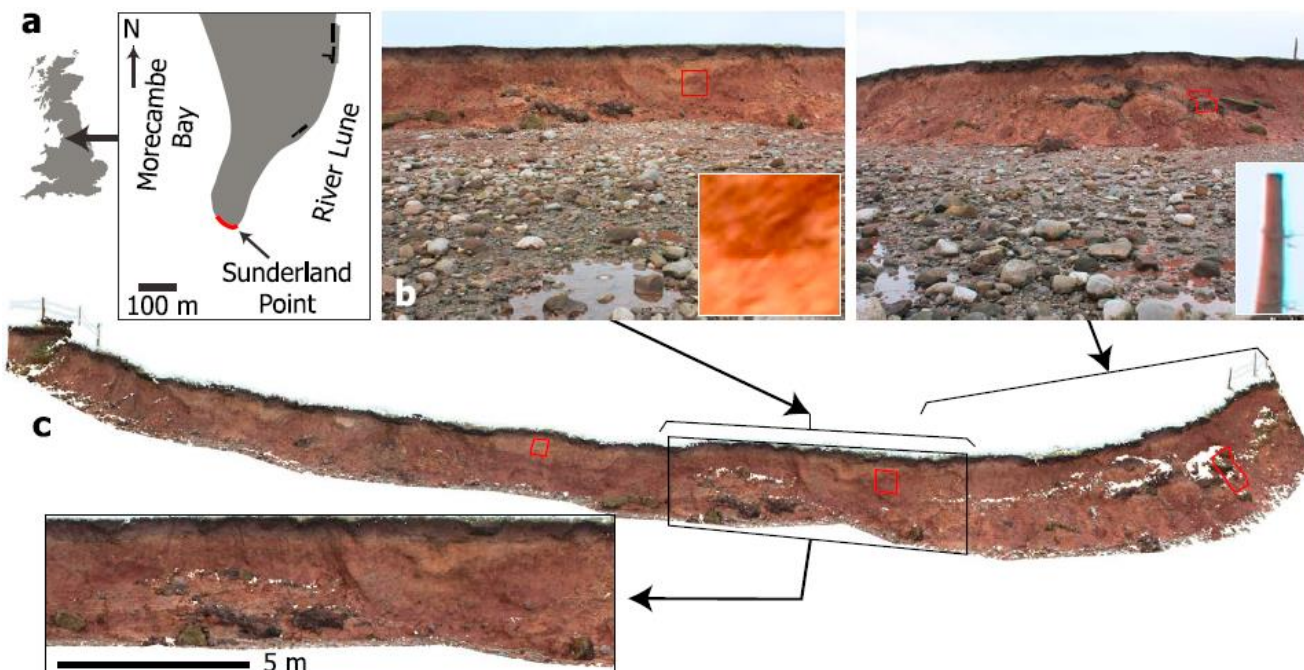
… reconstruct the scene **structure** (i.e. the geometry of the target *and* the positions, orientations & lens parameters of the cameras)

Where it all started…
First geoscience applications

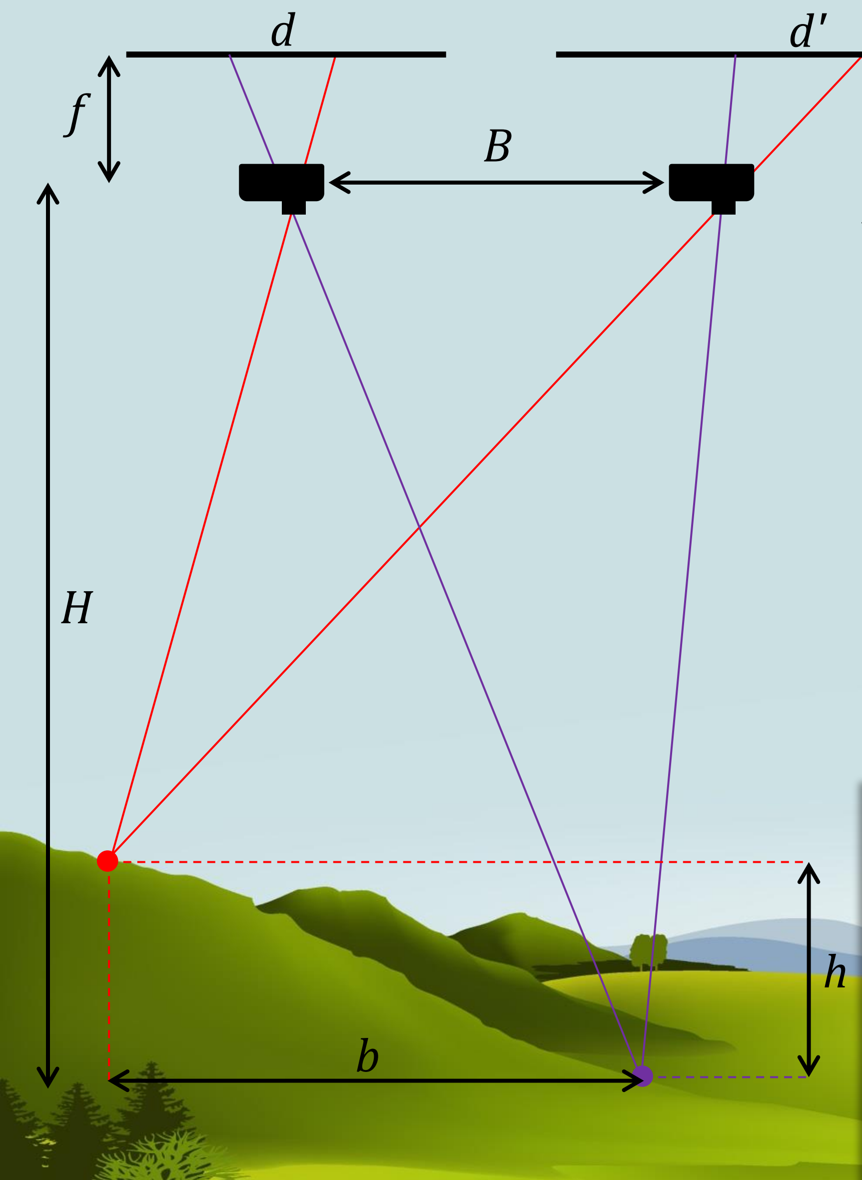


Left. Westoby *et al.* (2012). Structure-from-Motion photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology*

Right. James & Robson (2012). Straightforward reconstruction of 3D surfaces and topography with a camera: Accuracy and geoscience application. *Journal of Geophysical Research*



Traditional stereo-photogrammetry



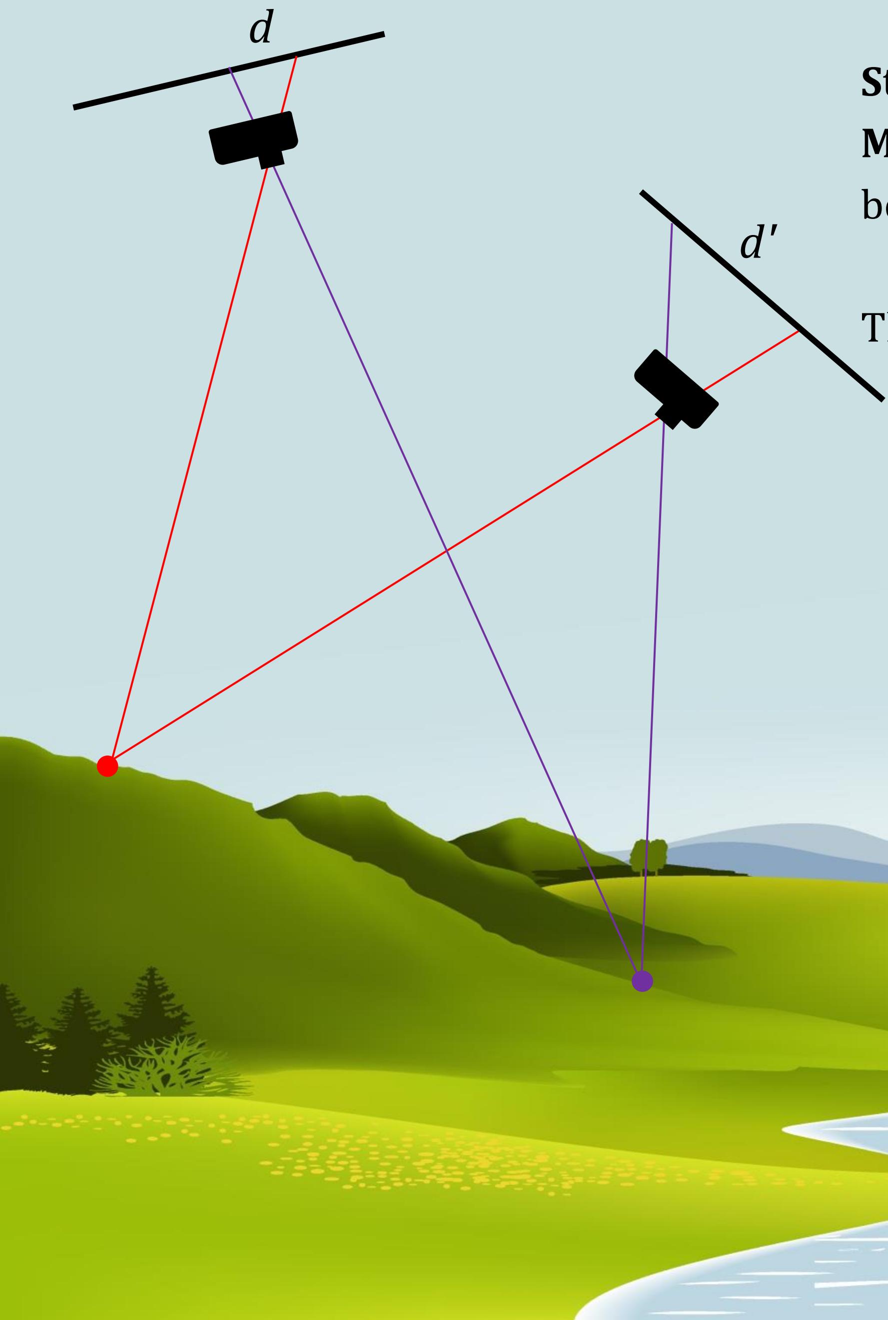
Known camera height *H* and focal length *f*, and the baseline *B* between images

Match corresponding features

Measure distances between features on the camera image plane *d, d'*

Calculate relative positions of features *b, h*

Structure-from-Motion



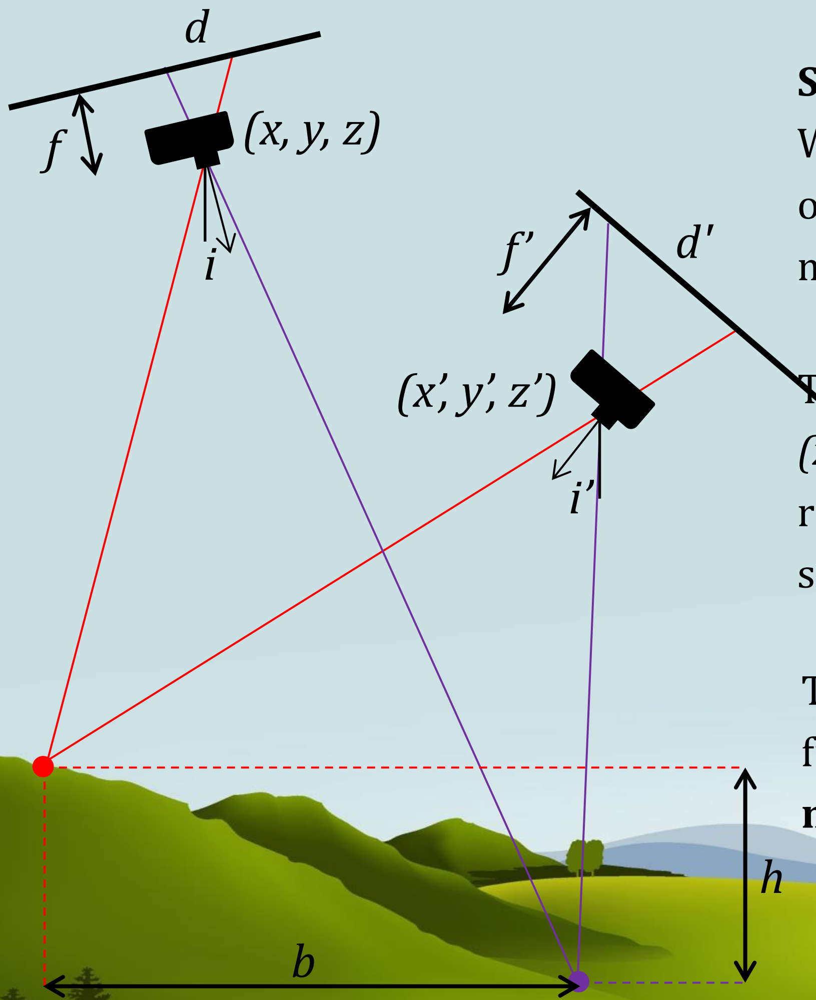
Step 1

Match corresponding features and measure distances between them on the camera image plane *d, d'*

The Scale Invariant Feature Transform is key to matching corresponding features despite varying distances

• The **Scale Invariant Feature Transform (SIFT)** (Lowe, 1999) allows corresponding features to be matched even with large variations in scale and viewpoint and under conditions of partial occlusion and changing illumination

Structure-from-Motion



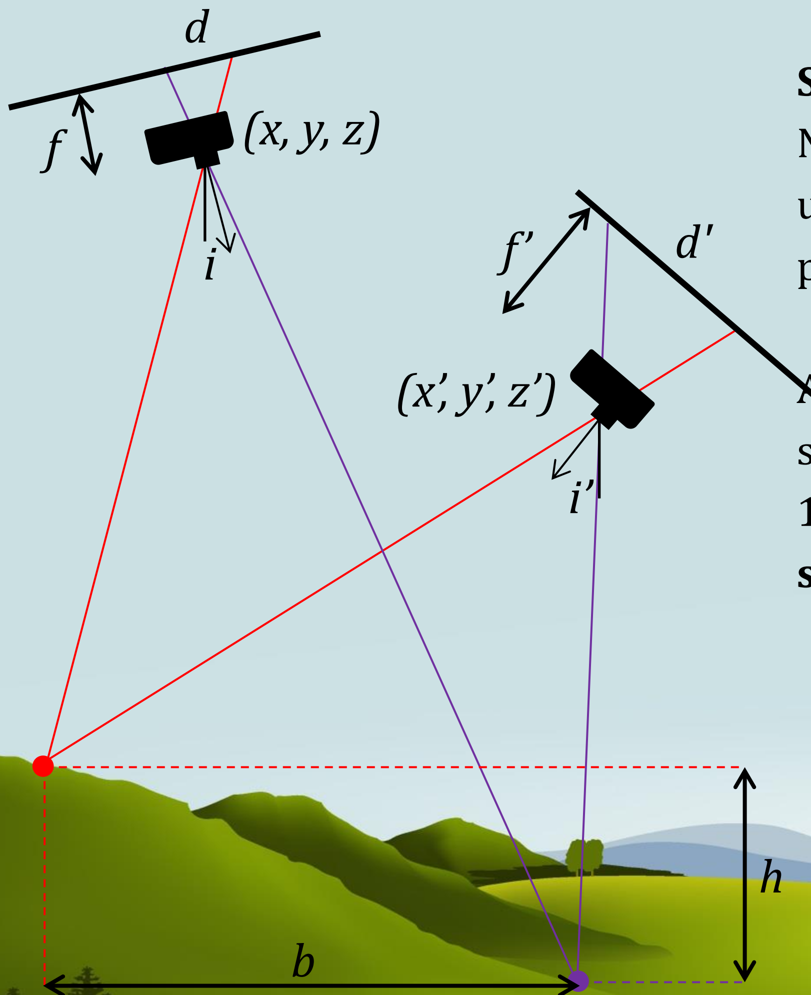
Step 2 Produce a "sparse cloud"

When we have the matching locations of multiple points on two or more photos, there is usually just one mathematical solution for where the photos were taken.

Therefore, we can calculate individual camera positions *(x, y, z), (x', y', z')*, orientations *i*, *i'*, focal lengths *f*, *f'*, and relative positions of corresponding features *b, h,* in a single step known as **"bundle adjustment".**

This is where the term Structure from Motion comes from. Scene **structure** refers to all these parameters; **motion** refers to movement of the camera

Structure-from-Motion

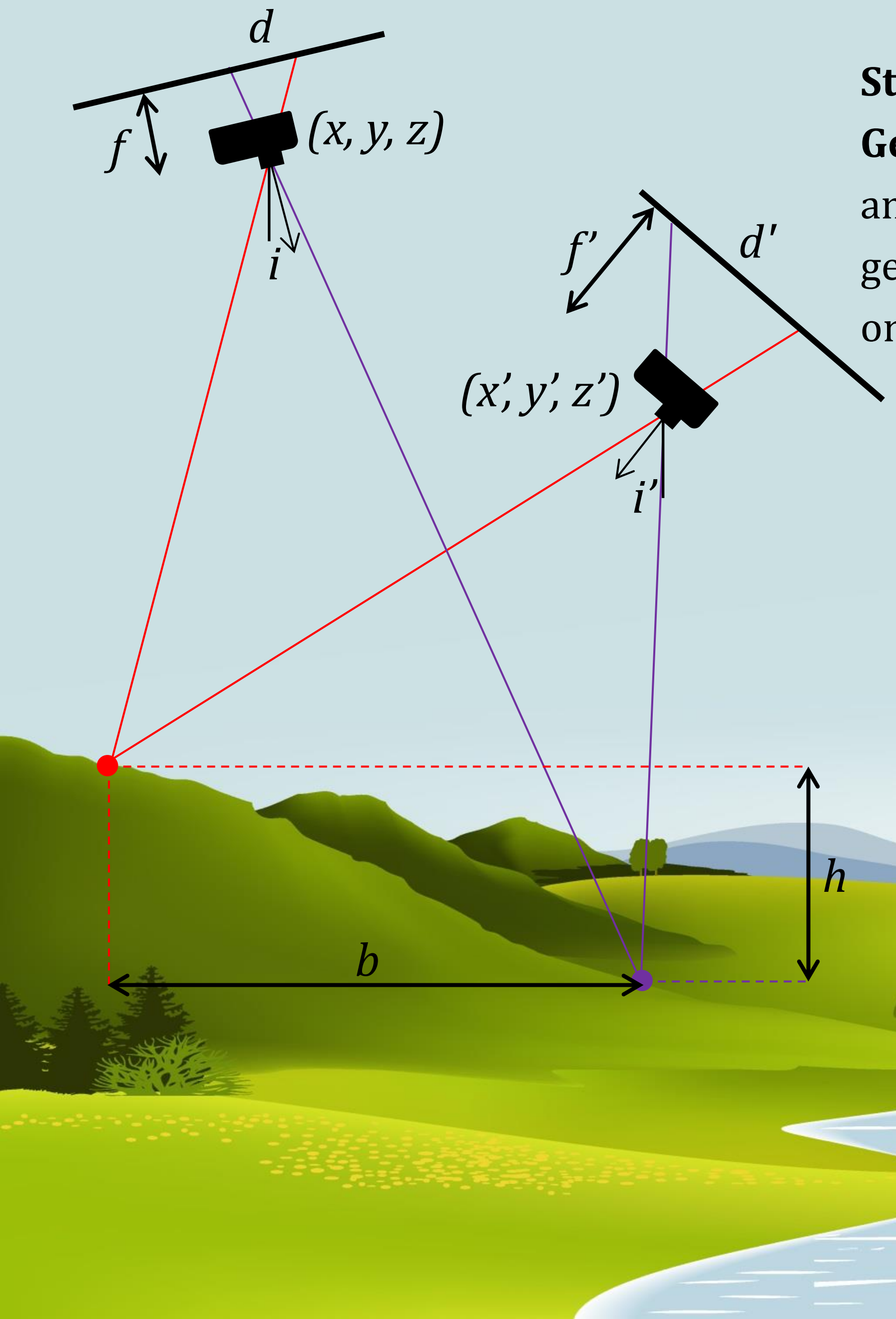


Step 3

Next, a dense point cloud and 3D surface is determined using the known camera parameters and using the SfM points as "ground control".

All pixels in all images are used so the dense model is similar in resolution to the raw photographs (typically 100s – 1000s point/m^2). This step is called "**multiview stereo matching" (MVS)**

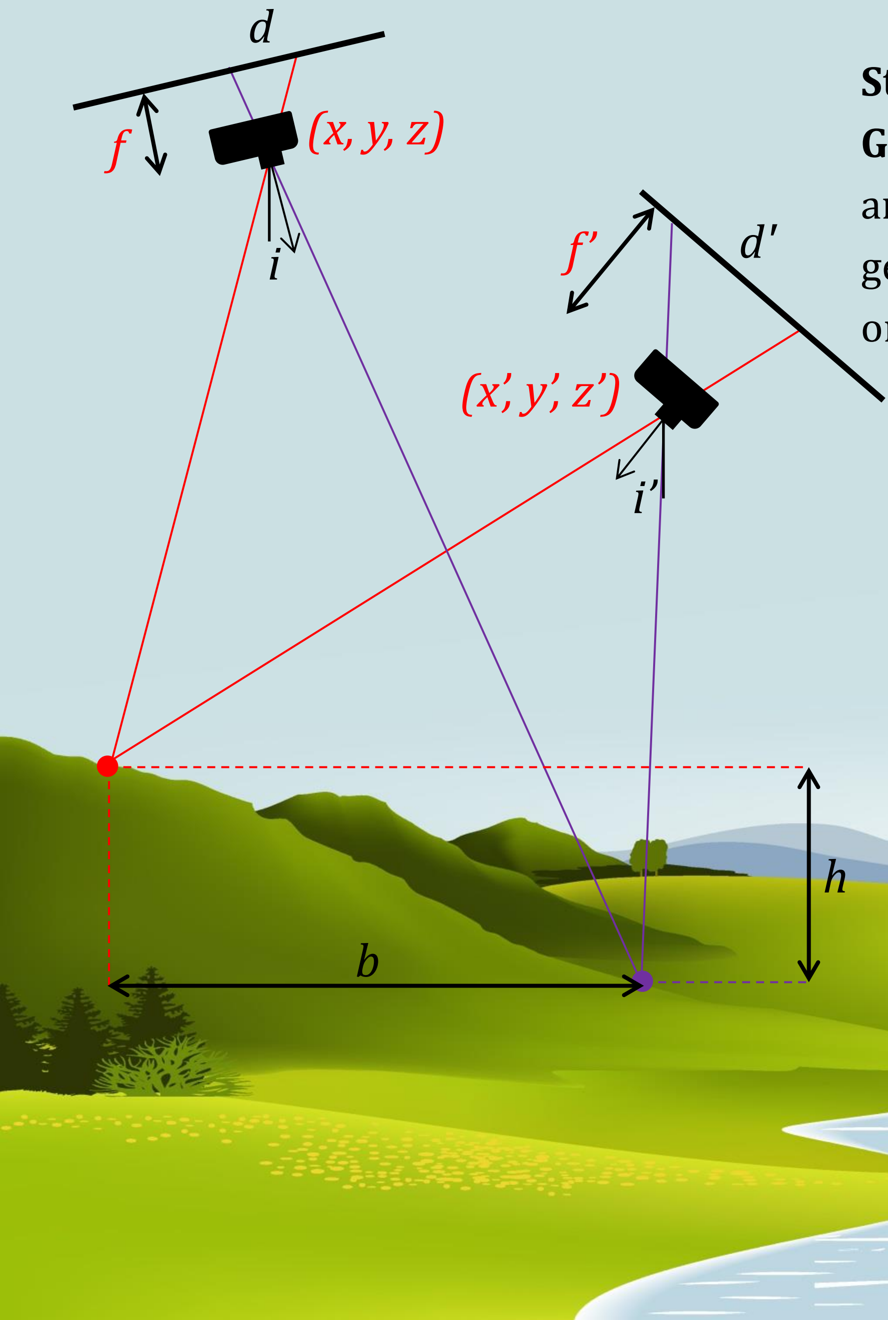
Structure-from-Motion



Step 4

Georectification means converting the point cloud from an internal, arbitrary coordinate system into a geographical coordinate system. This can be achieved in one of two ways:

Structure-from-Motion

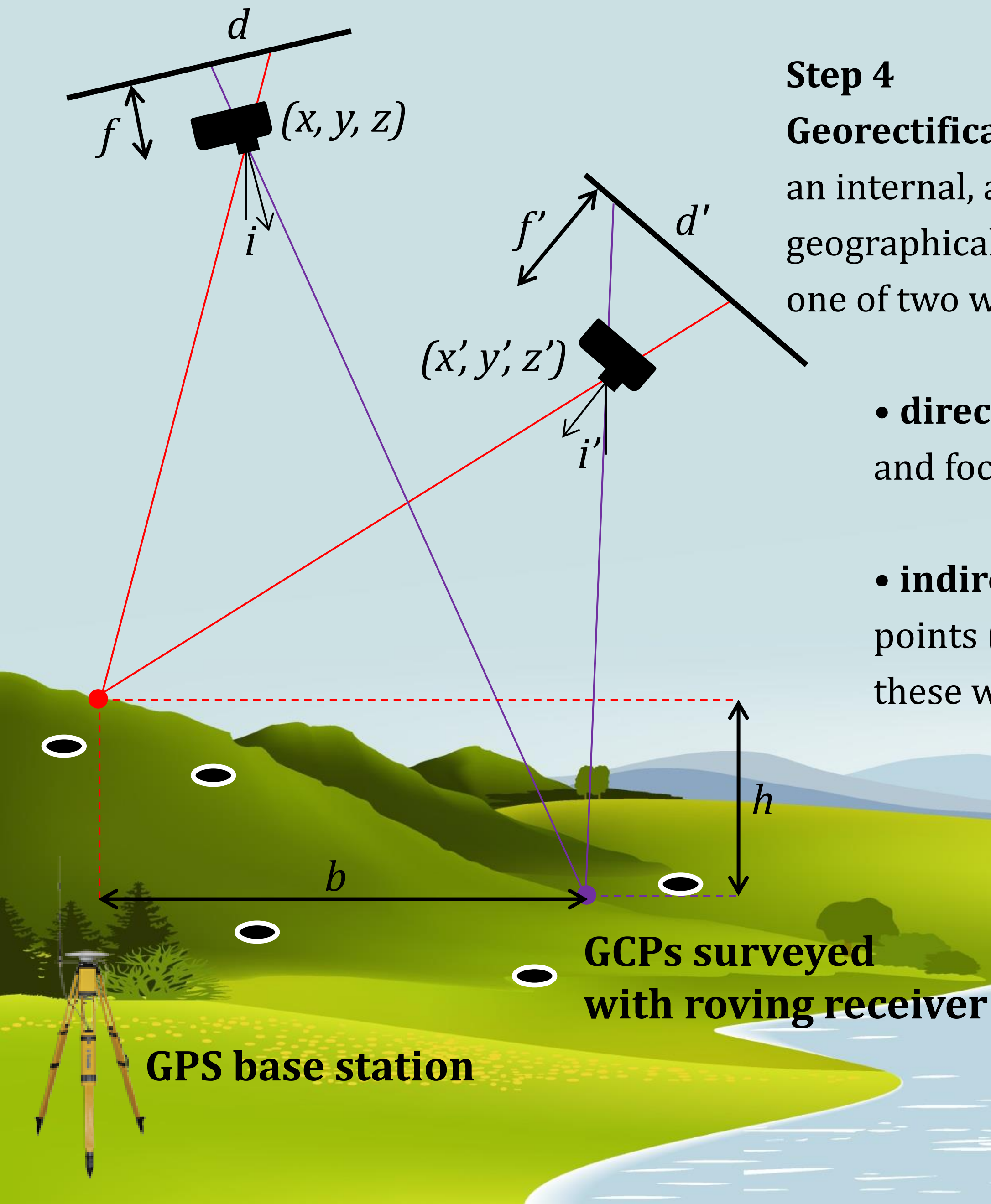


Step 4

Georectification means converting the point cloud from an internal, arbitrary coordinate system into a geographical coordinate system. This can be achieved in one of two ways:

- **directly,** with knowledge of the camera positions and focal lengths

Structure-from-Motion



Step 4

Georectification means converting the point cloud from an internal, arbitrary coordinate system into a geographical coordinate system. This can be achieved in one of two ways:

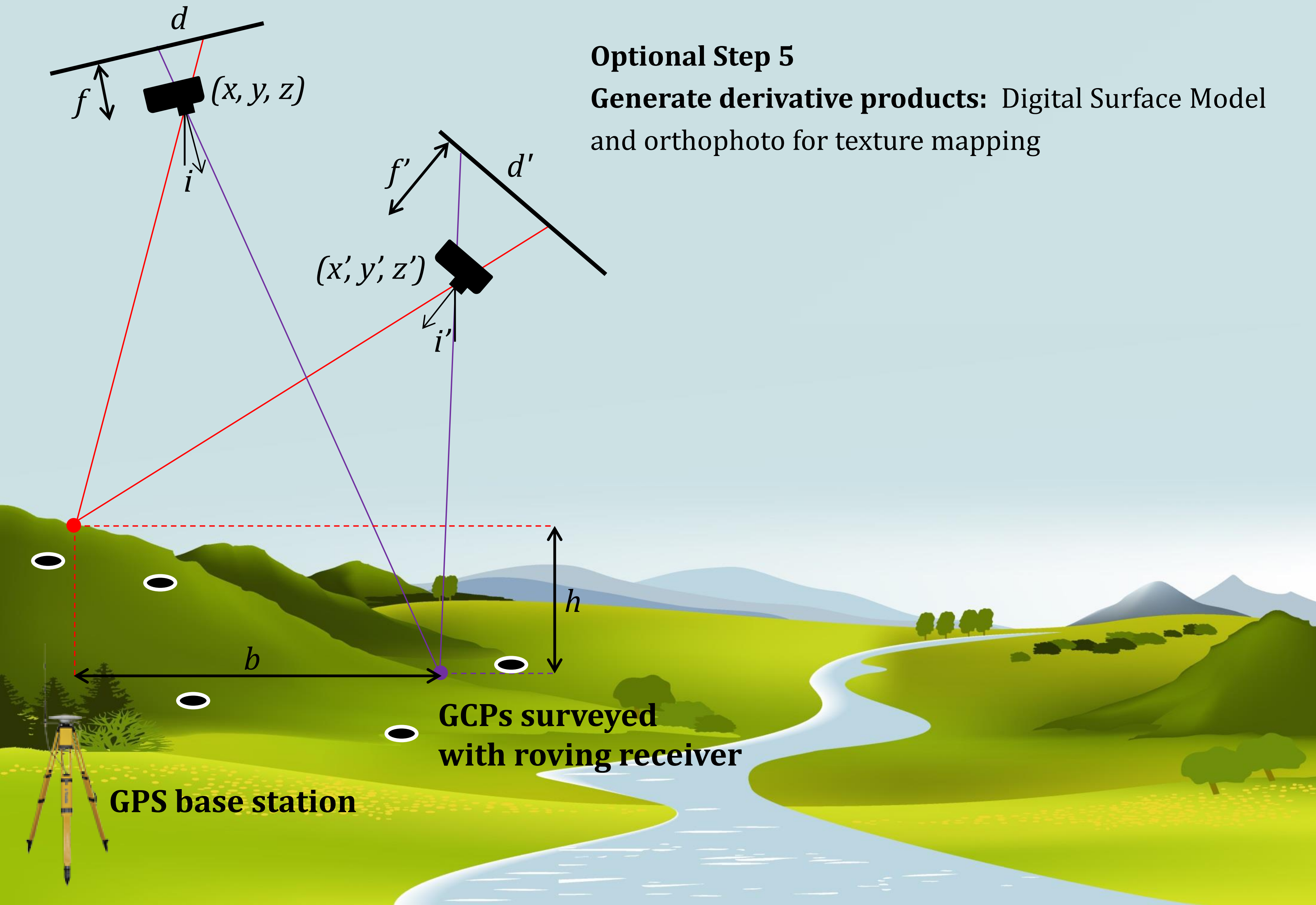
- **directly,** with knowledge of the camera positions and focal lengths

- **indirectly,** by incorporating a few ground control points (GCPs) with known coordinates. Typically these would be surveyed using differential GPS

Structure-from-Motion

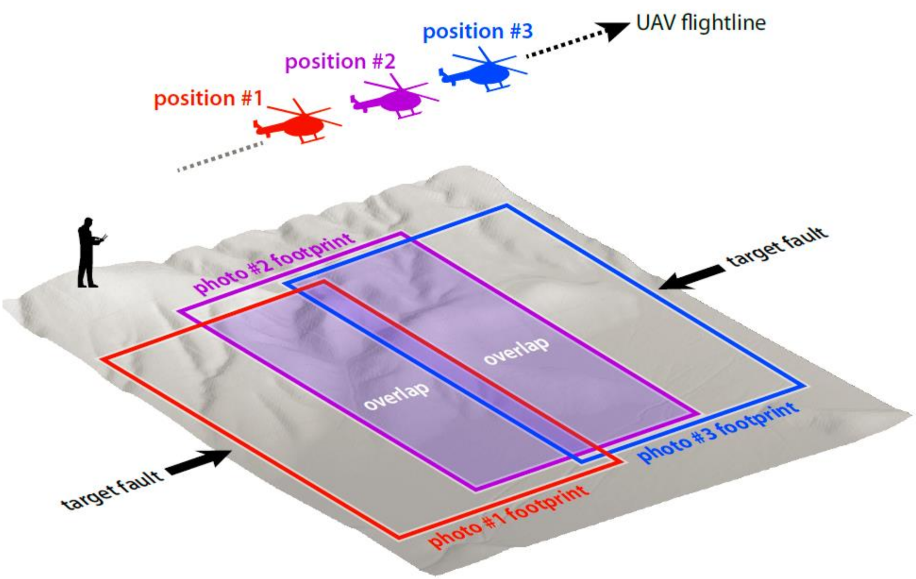
Optional Step 5

Generate derivative products: Digital Surface Model and orthophoto for texture mapping



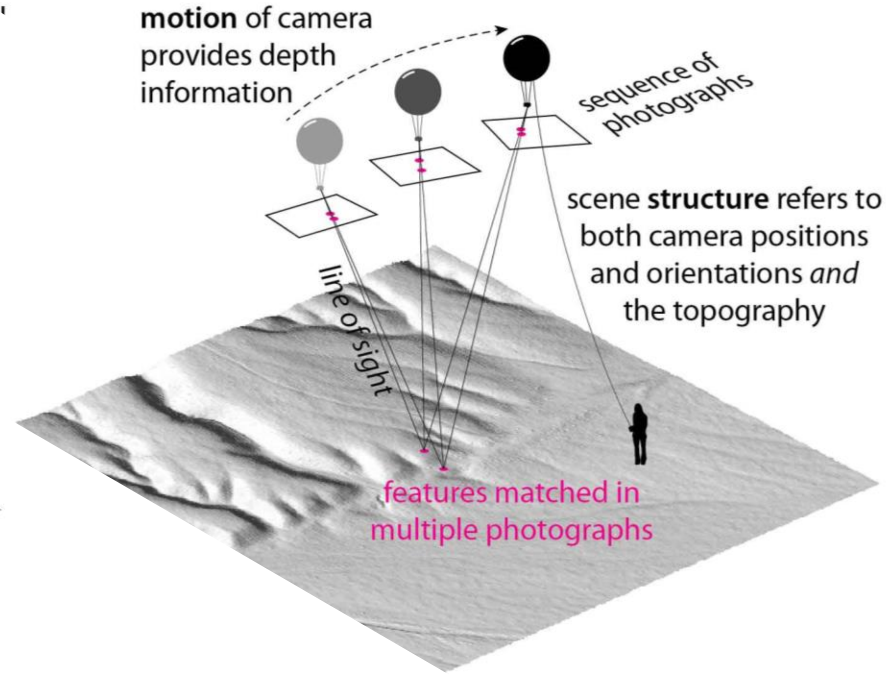
Traditional stereo-photogrammetry

- Requires a stable platform such as a satellite or aeroplane at a fixed elevation
- Photographs collected at known positions with fixed orientations and incidence angles



Structure-from-Motion

- Photos from many angles and distances can be used, with no *a priori* knowledge of locations or pose
- Enables "unstructured" image acquisition from the ground, legacy air-photosets, or unmanned platforms

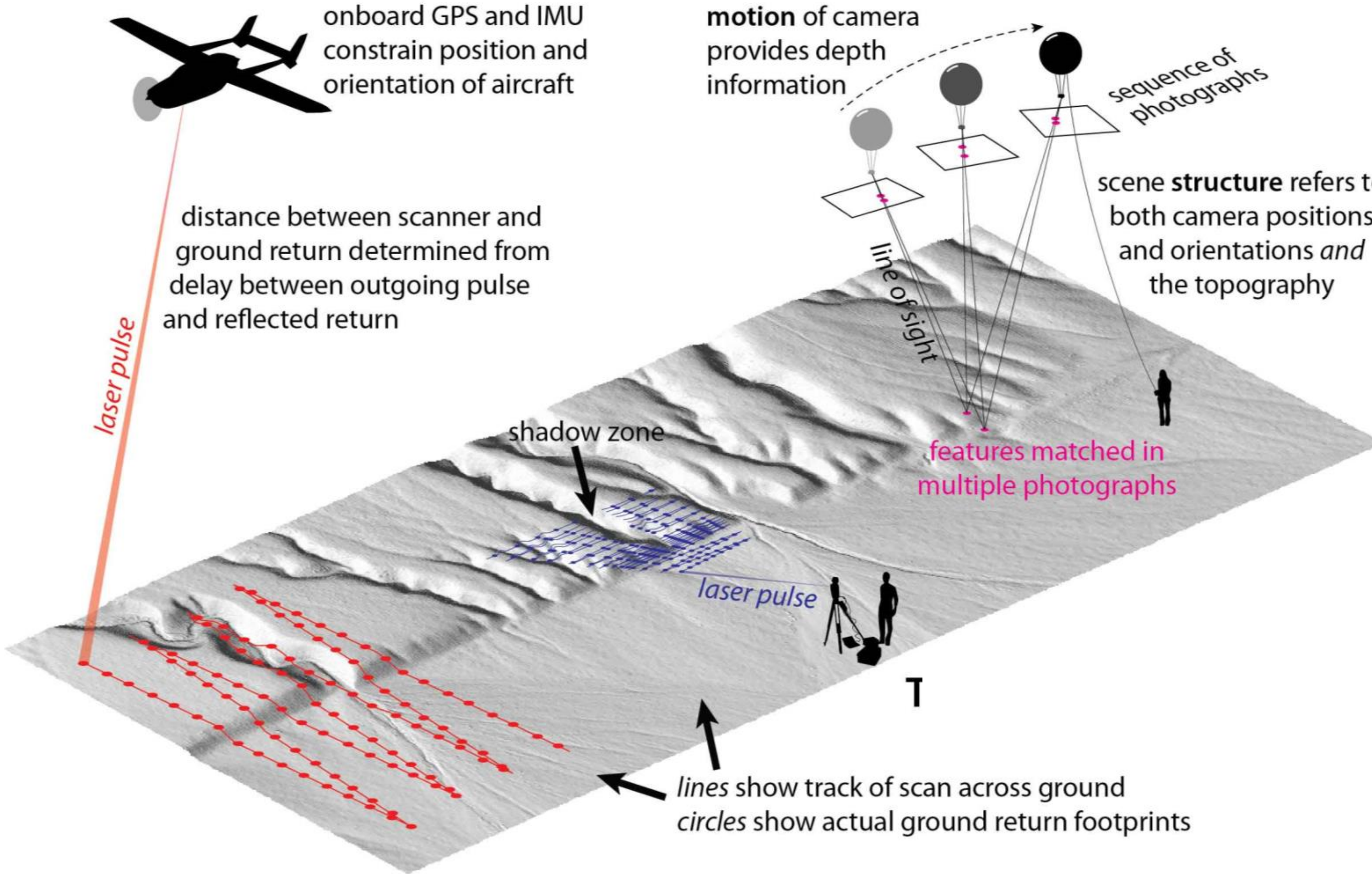


Lidar (ALS, TLS, MLS)

- Expensive laser equipment required
- Works in densely-vegetated landscapes
- Uses precise time-of-flight measurements but prone to artifacts from GPS and IMU

Structure-from-Motion

- Requires only a cheap camera
- Coloured points & orthophoto for texture mapping
- Back-solves for camera parameters; warping artifacts are a common problem but easily mitigated



SfM & MVS software

Table 1
Examples of open source and commercial software for photo-based 3d reconstruction.

Software	Url (valid on 17 May, 2014)	Notes
Freely available		
Bundler Photogrammetry Package[a,b]	http://blog.neonascent.net/archives/bundler-photogrammetry-package/	Used in James and Robson (2012). Script-based, no graphical user interface (GUI). Windows OS only.
SFMToolkit[a,b]	http://www.visual-experiments.com/demos/sfmtoolkit/	Similar software to above.
Python Photogrammetry Toolbox (PPT)[a,b]	http://code.google.com/p/osm-bundler/	Formerly OSM-bundler. Python-driven GUI and scripts, with a Linux distribution.
VisualSFM[b]	http://www.cs.washington.edu/homes/ccwu/vsfm/	Advanced GUI with Windows, Linux and Mac. OSX versions. Georeferencing options, but camera model is more restricted than that used in Bundler.
3DF Samantha	http://www.3dflow.net/technology/samantha-structure-from-motion/	SfM only, but with more advanced camera models than all above (Farenzena et al., 2009). Provides output compatible with several dense matching algorithms.
http://opendronemap.org/		
Web sites and services		
Photosynth	http://photosynth.net/	Evolved from Bundler. SfM only, no dense reconstruction. Can incorporate a very wide variety of images, but does so at the cost of reconstruction accuracy.
Arc3D	http://www.arc3d.be/	Vergauwen and Van Gool [2006]
CMP SfM Web service[a]	http://ptak.felk.cvut.cz/sfmservice/	
Autodesk 123D Catch	http://www.123dapp.com/catch/	
Pix4D	http://pix4d.com/	Also available as standalone software.
My3DScanner	http://www.my3dscanner.com/	
Commercial		
PhotoScan	http://www.agisoft.ru/products/photoscan/	Full SfM-MVS-based commercial package.
Acute3D	http://www.acute3d.com/	
PhotoModeler	http://www.photomodeler.com/	Software, originally based on close-range photogrammetry, now also implements SfM.
3DF Zephyr Pro	http://www.3dflow.net/	Underlying SfM engine is 3DF Samantha

Note: Table modified from http://www.lancaster.ac.uk/staff/jamesm/research/sfm.htm.
SfM = Structure from Motion; MVS = Multi-View Stereo.
[a] Uses Bundler (http://phototour.cs.washington.edu/bundler/) to compute structure from motion.
[b] Uses PMVS2 (http://grail.cs.washington.edu/software/pmvs/) as a dense multi-view matcher.

Bemis *et al.* (2014). Ground-based and UAV-Based photogrammetry: A multi-scale, high resolution mapping tool for structural geology and paleoseismology. *Journal of Structural Geology*

SfM from Unmanned Aerial Vehicles (UAV)


DJI Phantom 2 quadcopter (~$1k)


Custom built helicopter (~$15k)


Autokite (~$1k, discontinued)

Falcon Unmanned fixed wing (~$12k)

SfM from helicopters and multi-rotor UAVs

Pros Robust in high wind and can take off and land anywhere. Larger helicopters can carry large SLR camera. Smaller multi-rotors cannot, but are easier to fly.

Cons Helicopter needs trained pilot to take-off and land and regular refuelling. Initial costs are high and requires careful maintenance.

Regulations may need to be followed (FAA in the U.S.)

SfM from fixed wing UAVs

Pros Relatively easy to pilot. Can cope in moderate winds. Flight durations are normally longer than copters.

Cons Susceptible to damage during landing.

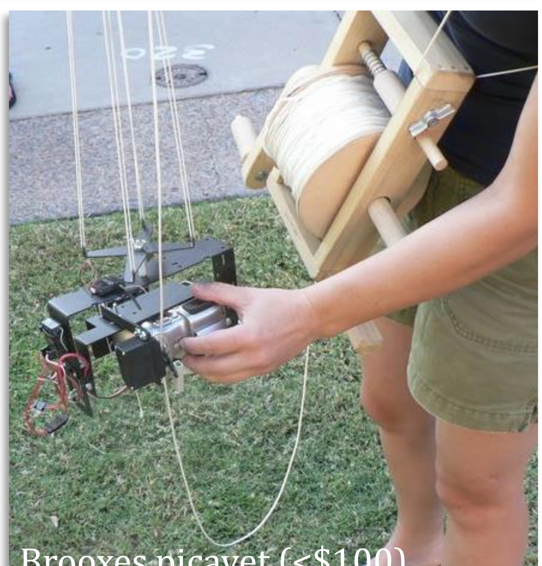
Regulations may need to be followed (FAA in the U.S.)

Autokite (~$1k, discontinued)

Falcon Unmanned fixed wing (~$12k)

SfM from Unmanned Aerial Systems (UAS)

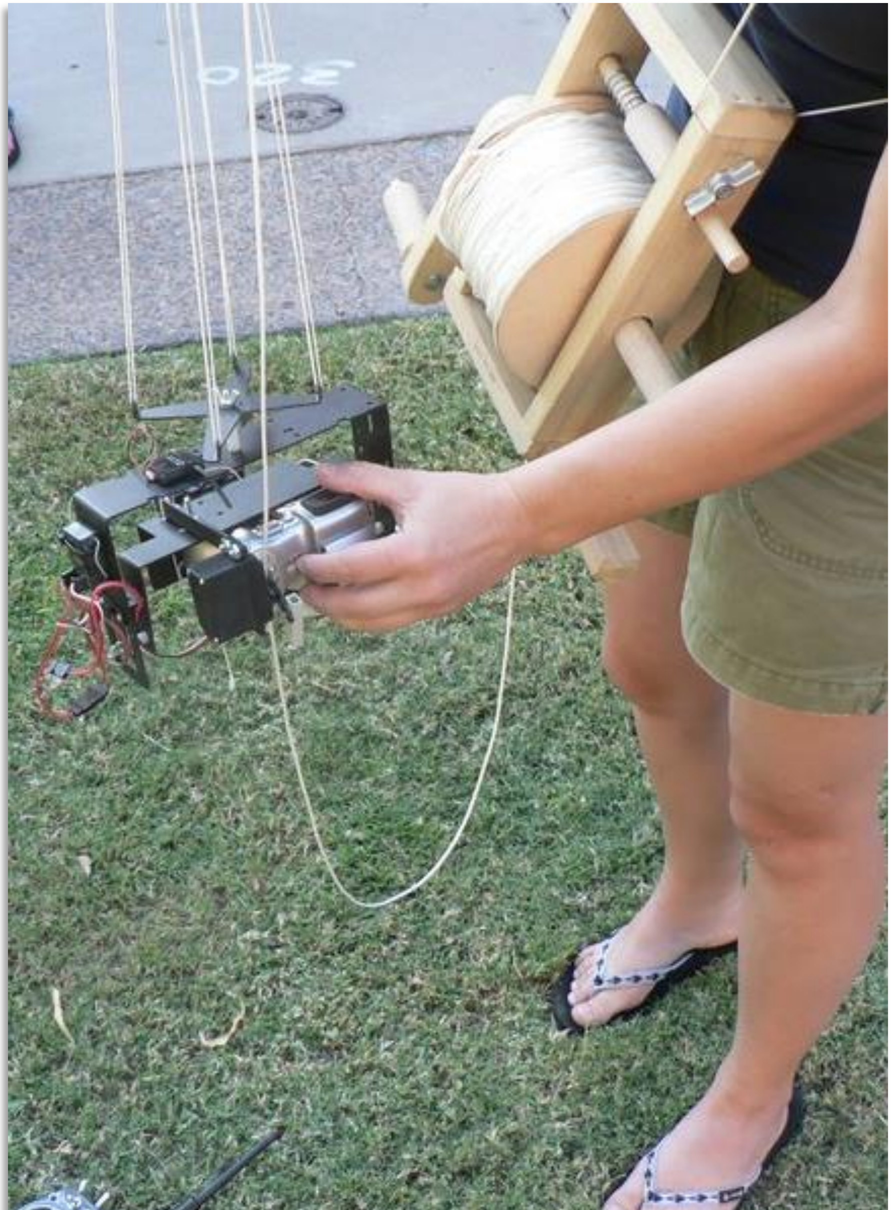
Allsopp helikite (~$2k)



Brooxes picavet (<$100)

Ramon's balloon (~$100s)

SfM from Unmanned Aerial Systems (UAS)



Pros Easy to drag across target area. Once in the air can remain there. Can carry large SLR cameras. No FAA regulations!

Cons Requires helium, which can be expensive (>$100 per canister), and fiddly picavet. Cannot be automated. Difficult to deploy in windy conditions.

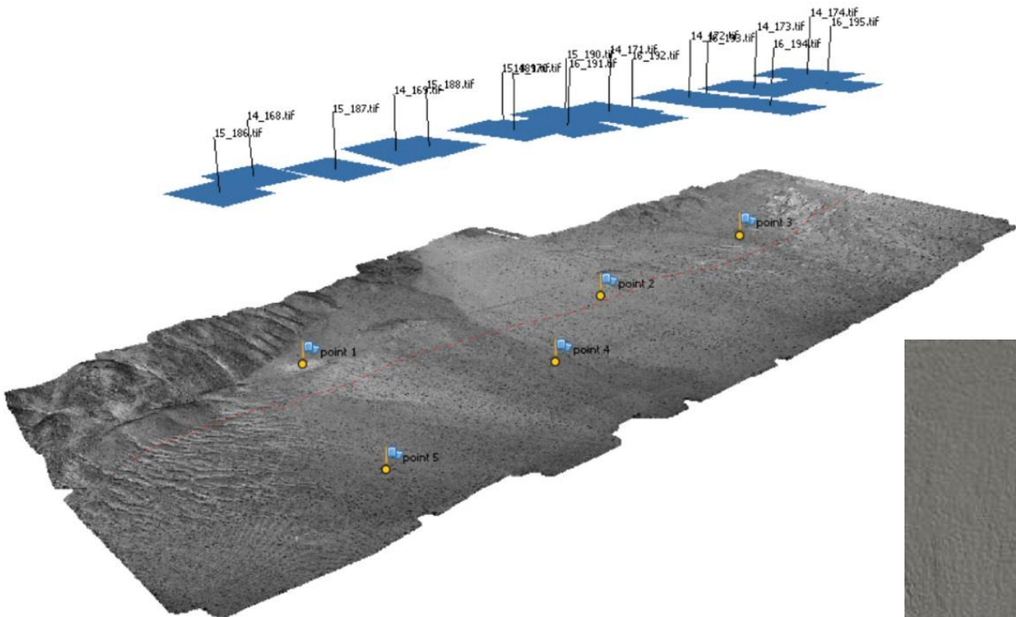
SfM from Unmanned Aerial Systems (UAS)

Pros Easy to drag across target area. Once in the air can remain there. Robust in high wind. No FAA regulations!

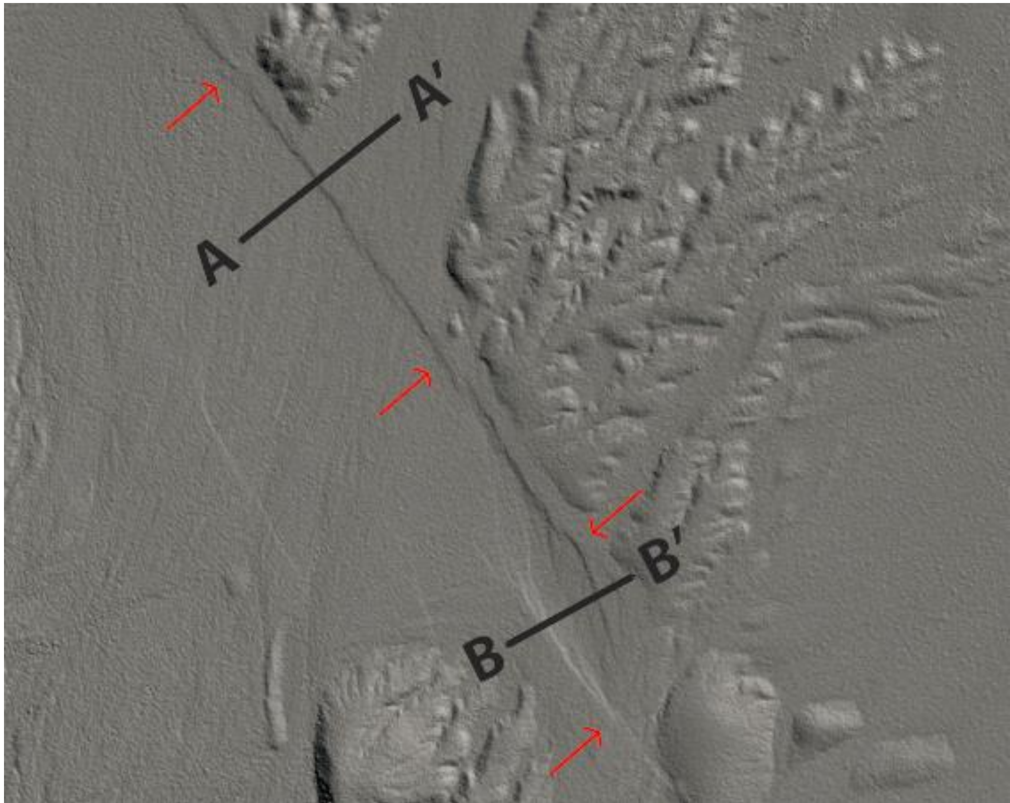
Cons Requires helium, which can be expensive (>$100 per canister). Cannot be automated. Carries small cameras.

SfM from airplane photos

• "Historical topography" and "diachronic geomorphology" possible using legacy air-photos. Requires sufficient photo overlap and georeferencing is a challenge.



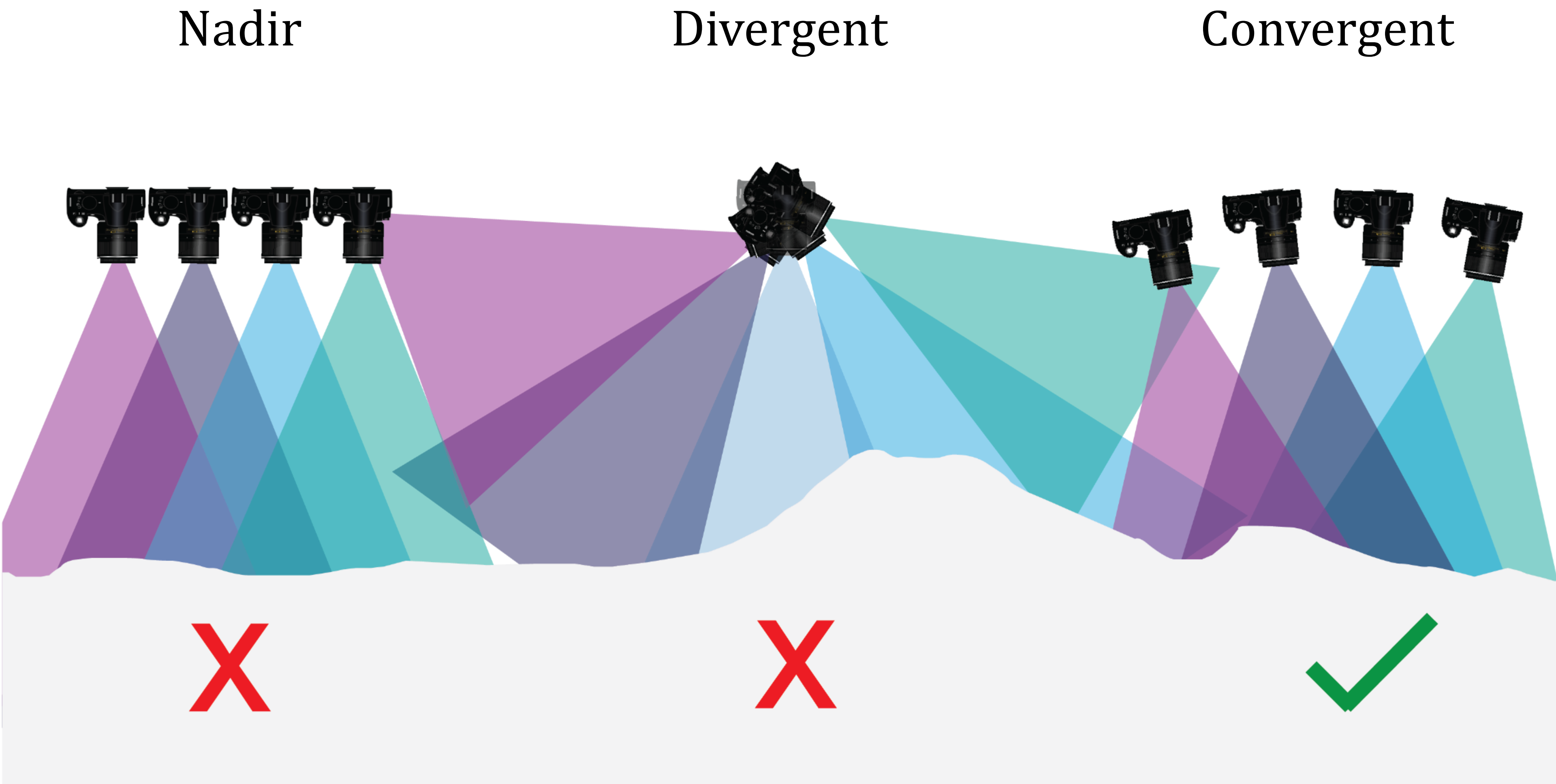
(Left) A short section of the ~85 km-long USGS aerial survey of the 1992 Landers rupture, California.



(Right) Resulting 30 cm-resolution DEM, hillshaded to highlight fine geomorphic features.

Georeferencing was undertaken using modern satellite imagery

Acquisition geometry



Acquisition geometry

Convergent with a range of distances

Choice of camera

- Most cameras work
- DEM/orthophoto resolution is governed by the ground pixel resolution of the raw photos, so high megapixel cameras are preferable
- Better lenses of SLR cameras mean fewer radial distortions…
- but radial artefacts arising from cheap camera lenses can be mitigated by deploying ground control points
- fish-eye lenses (e.g. GoPro) give rise to largest distortions, but latest software seems to cope
- **time lapse setting** is essential if camera is deployed from drone
- internal or external **GPS tagging** is another useful function, as it enables rough geo-referencing without ground control points

Camera lens distortions

f = focal length

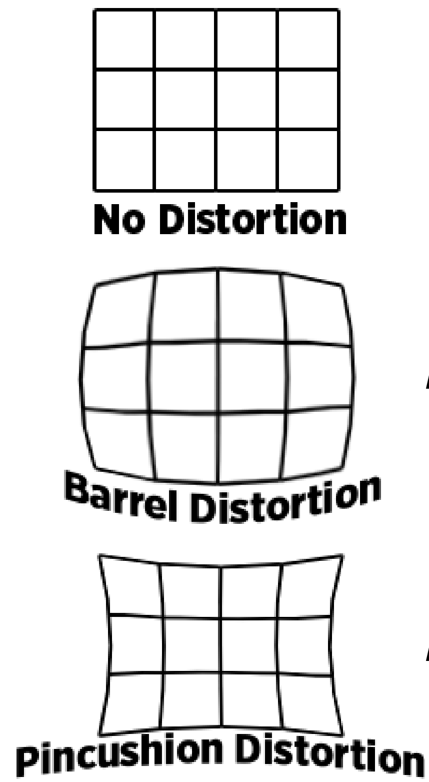
c_x = principal point x coordinate

c_y = principal point y coordinate

k_n = n^{th} radial distortion coefficient

p_n = n^{th} tangential distortion coefficient

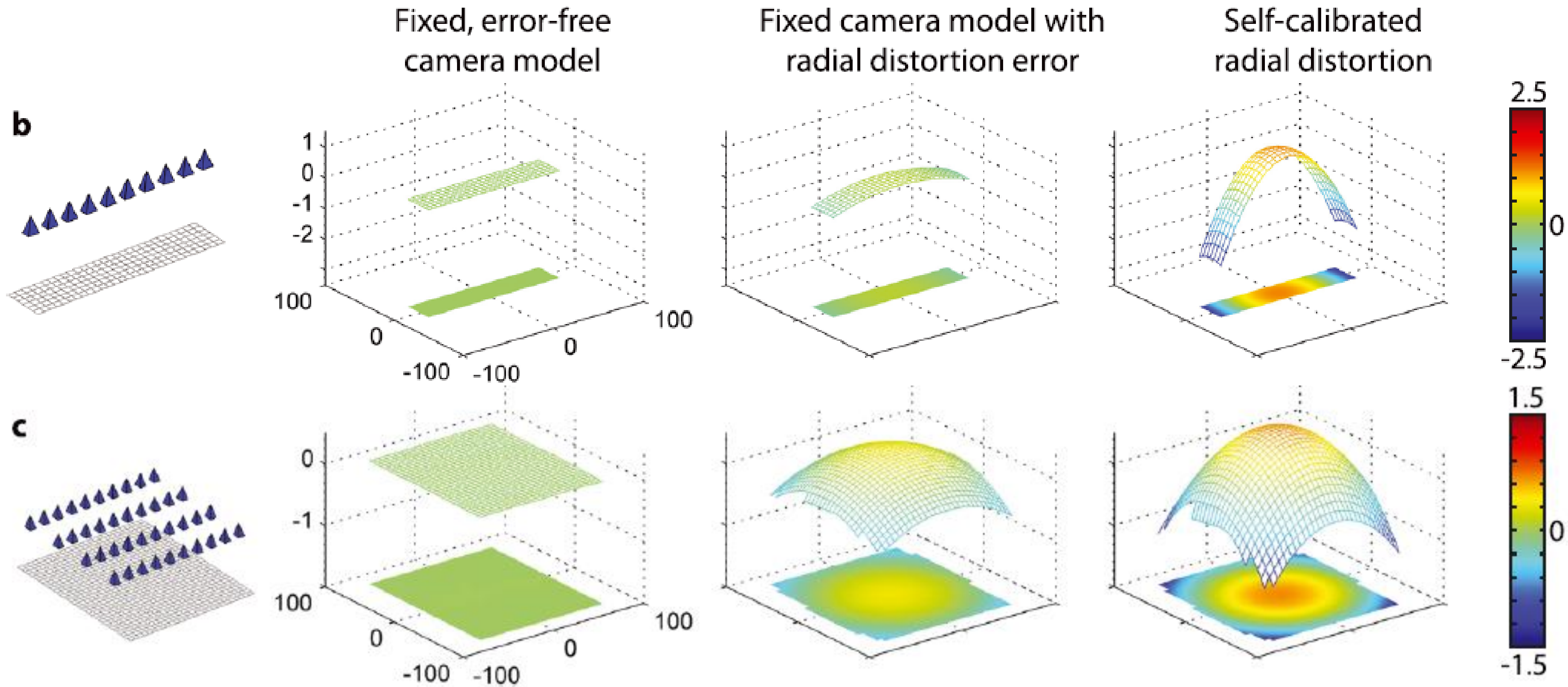
skew coefficient between the x and the y axis.



$k_1 < 1$

$k_1 > 1$

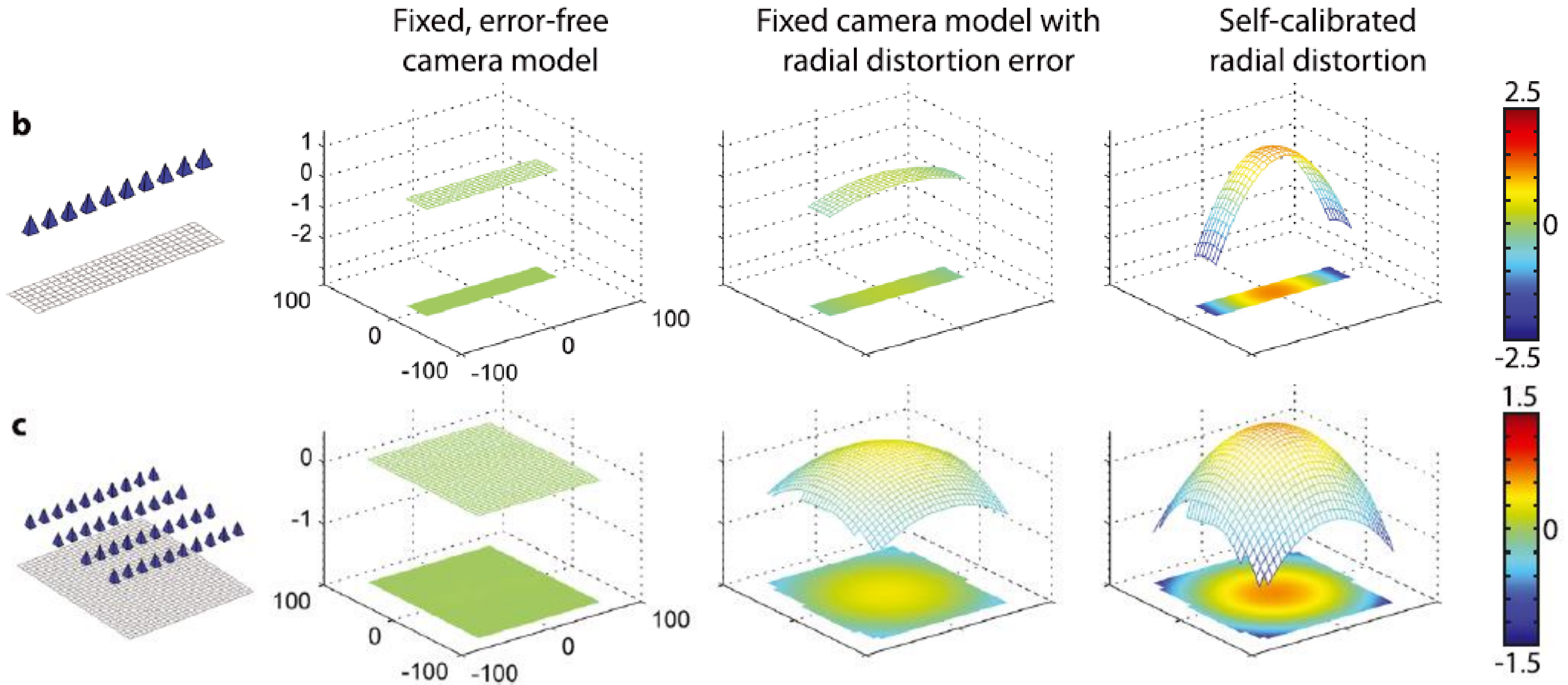
Camera lens distortions



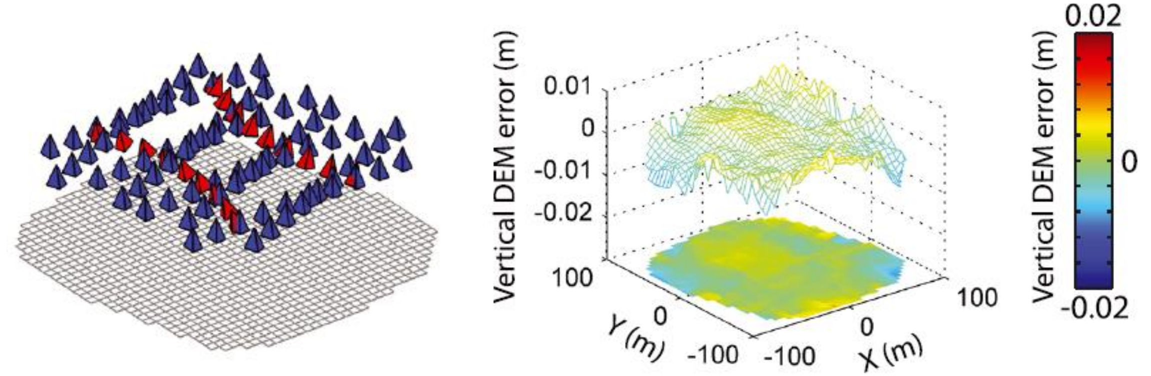
- A trade-off between lens radial distortion term and computed surface form can lead to "doming"

James & Robson (2014), Mitigating systematic error in topographic models derived from UAV and ground-based image networks, *Earth Surface Processes and Landforms*

Camera lens distortions



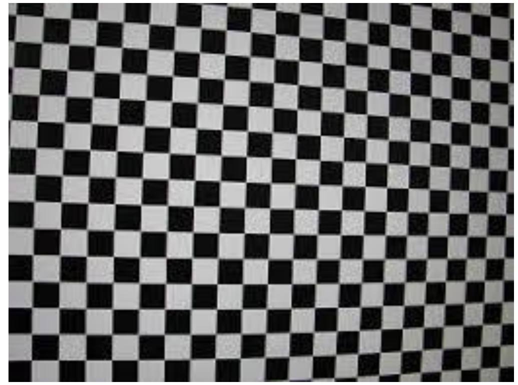
- Doming can be mitigated by incorporating a few oblique camera angles (in red)

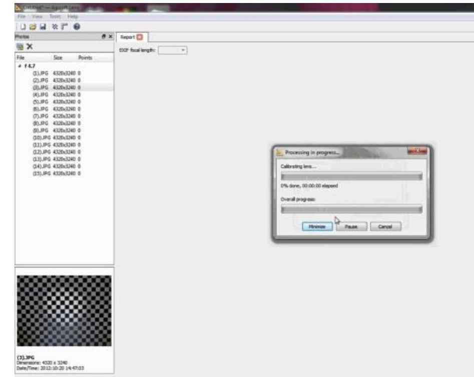


James & Robson (2014), Mitigating systematic error in topographic models derived from UAV and ground-based image networks, *Earth Surface Processes and Landforms*

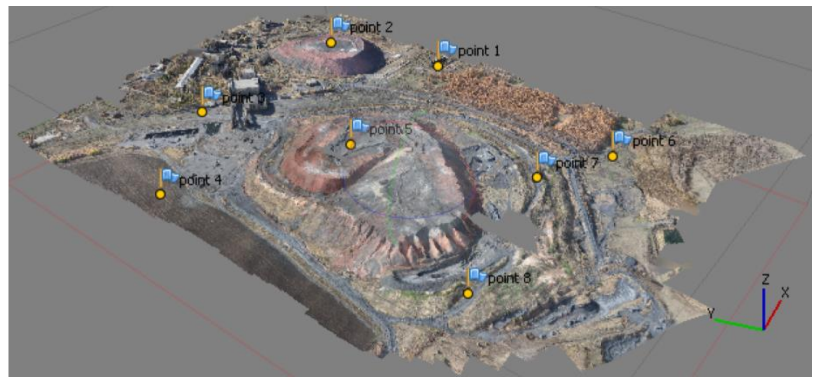
Camera lens distortions

• Doming can be mitigated by calibrating the camera parameters by photographing a calibration target

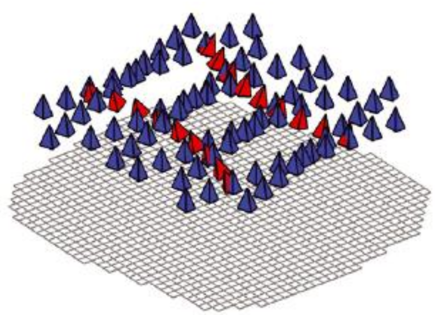


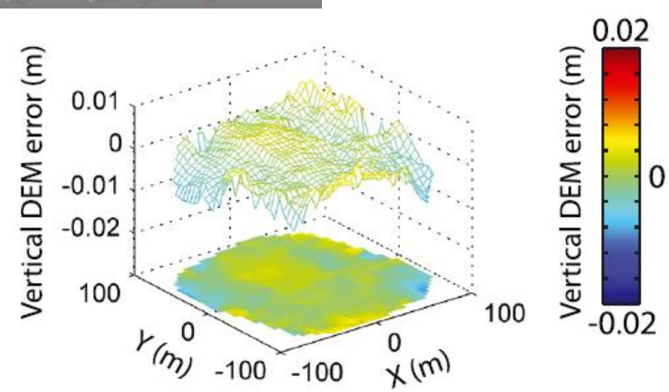


• Doming can be mitigated by georeferencing using ground control points

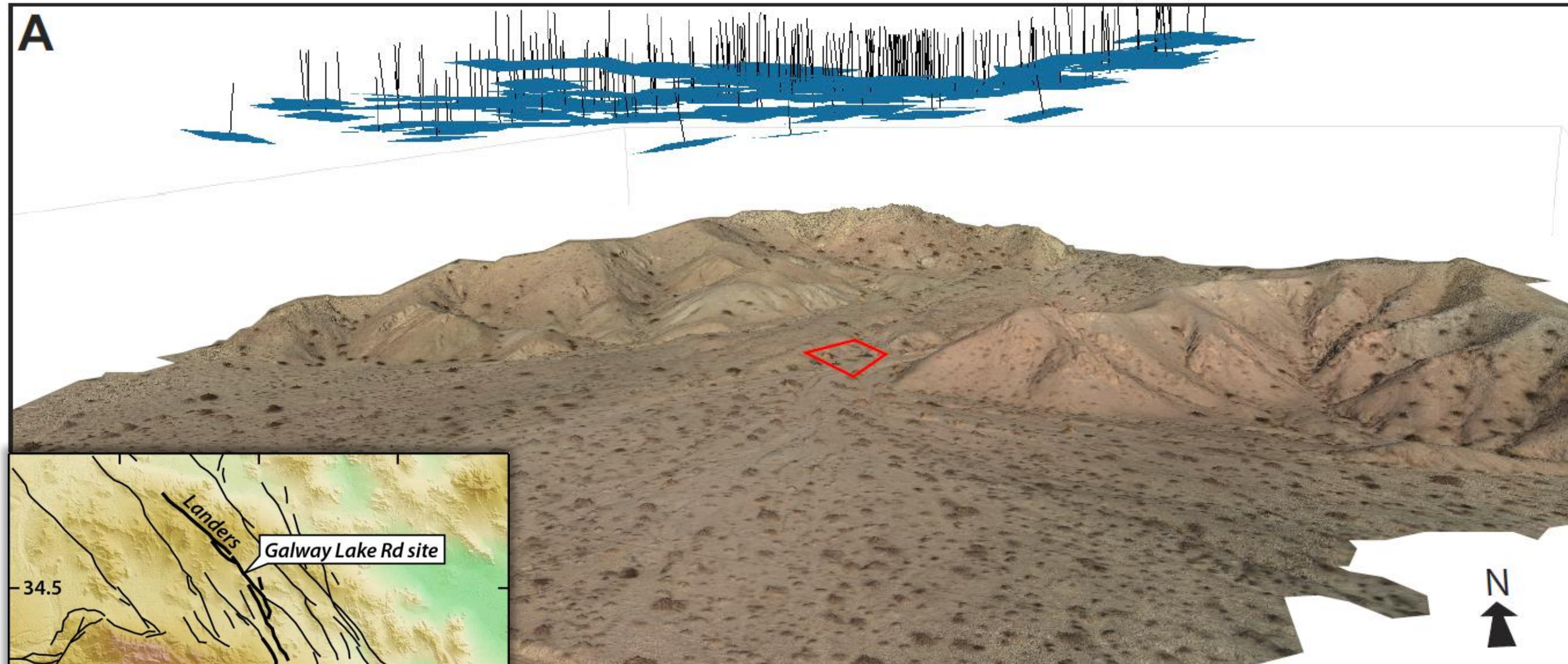


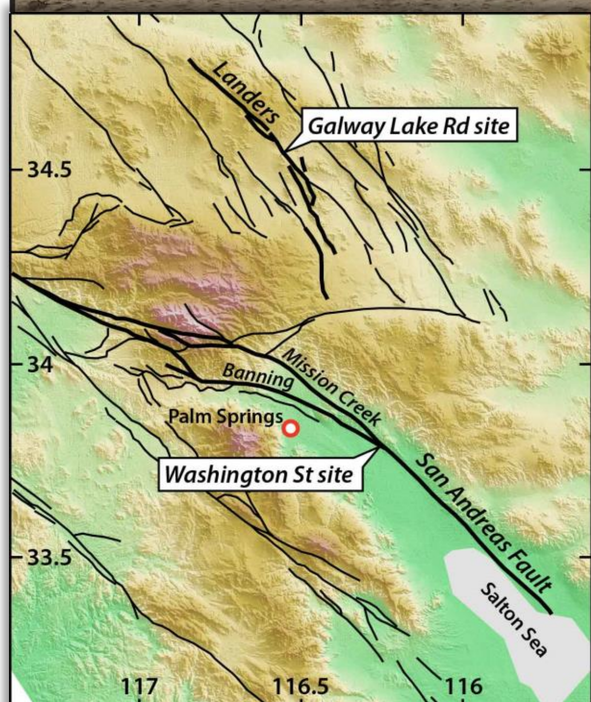
• Doming can be mitigated by incorporating a few oblique camera angles (in red)





Resolution and precision of SfM topography





It is important to capture each part of the target or target area with photos taken from several different locations. There needs to be significant overlap between images.

This image shows a test area in California where we made comparisons between SfM topography and airborne lidar. We used 230 photos taken in ~1 hour from a helium balloon.

Johnson *et al.* (2014), Rapid mapping of ultrafine fault zone topography with structure from motion, *Geosphere*

Resolution and precision of SfM topography

Orthophoto

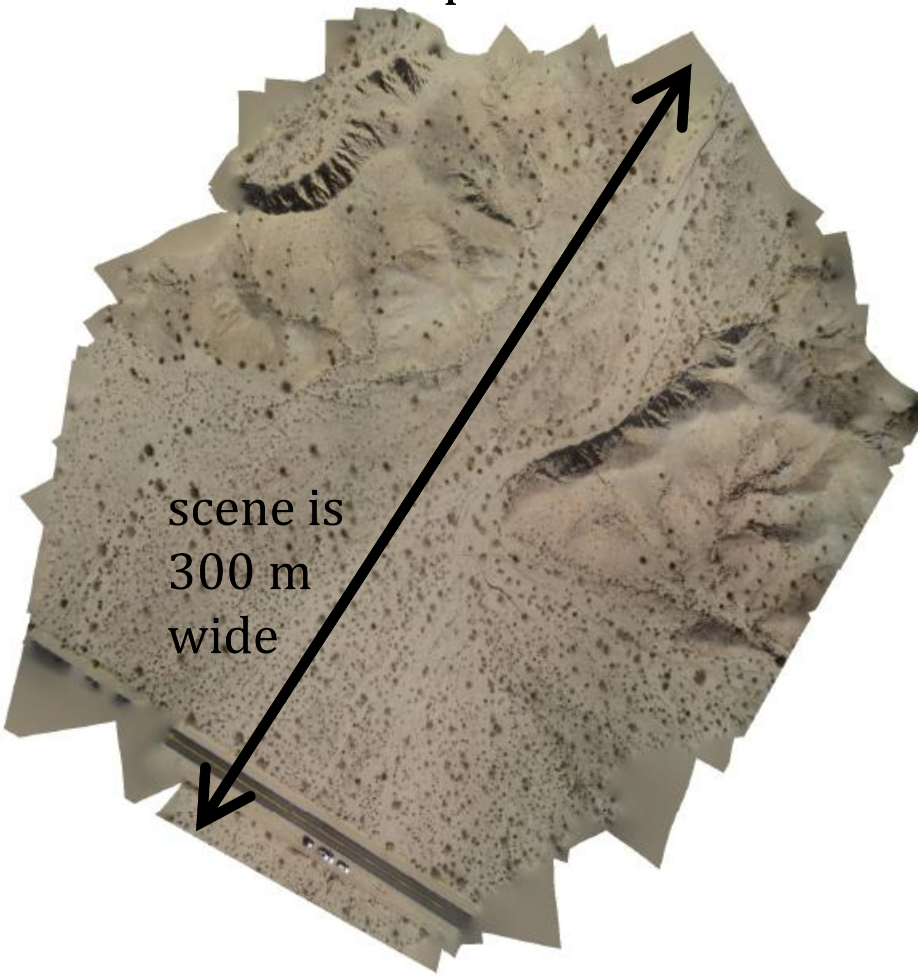


Photo coverage plot

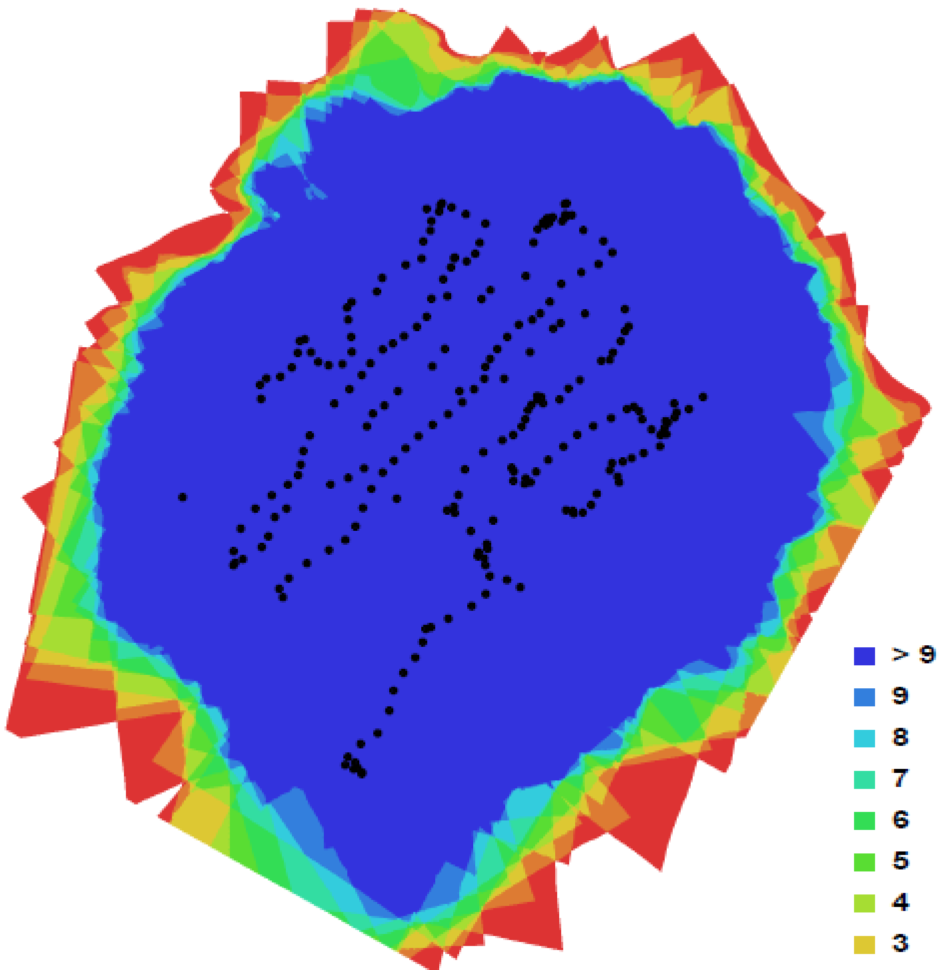
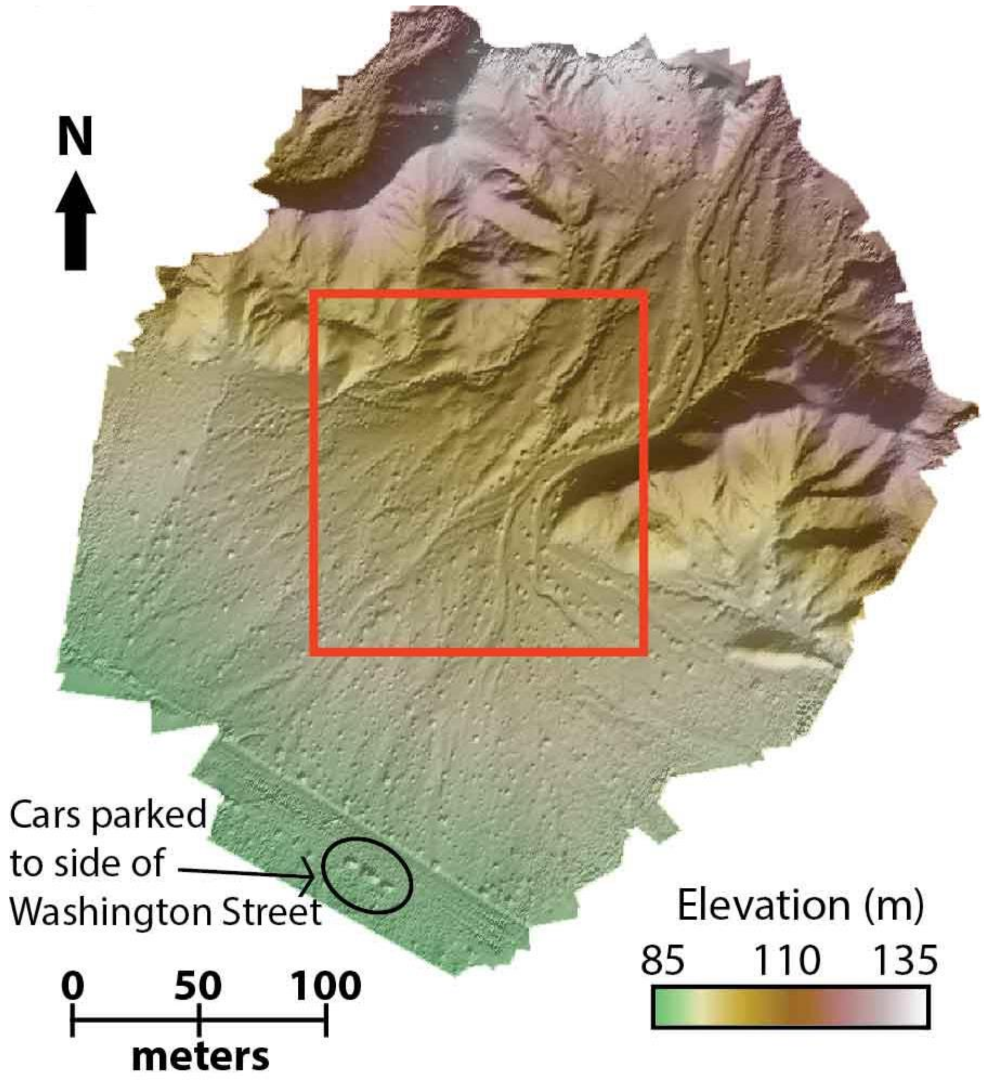
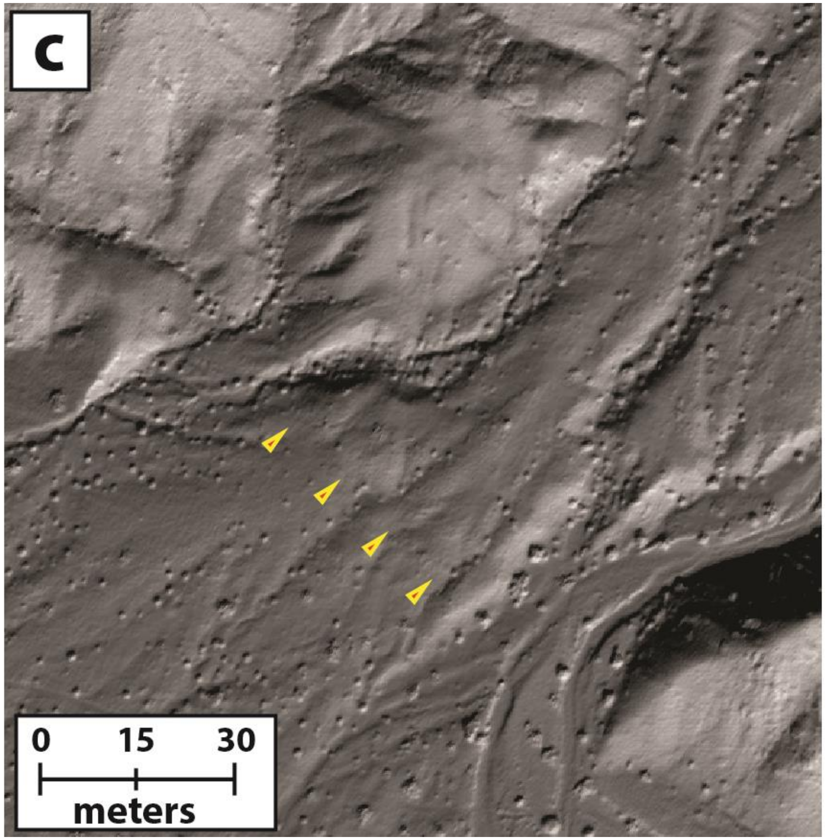


Fig. 1. Camera locations and image overlap.

Johnson *et al.* (2014), Rapid mapping of ultrafine fault zone topography with structure from motion, *Geosphere*

Resolution and precision of SfM topography



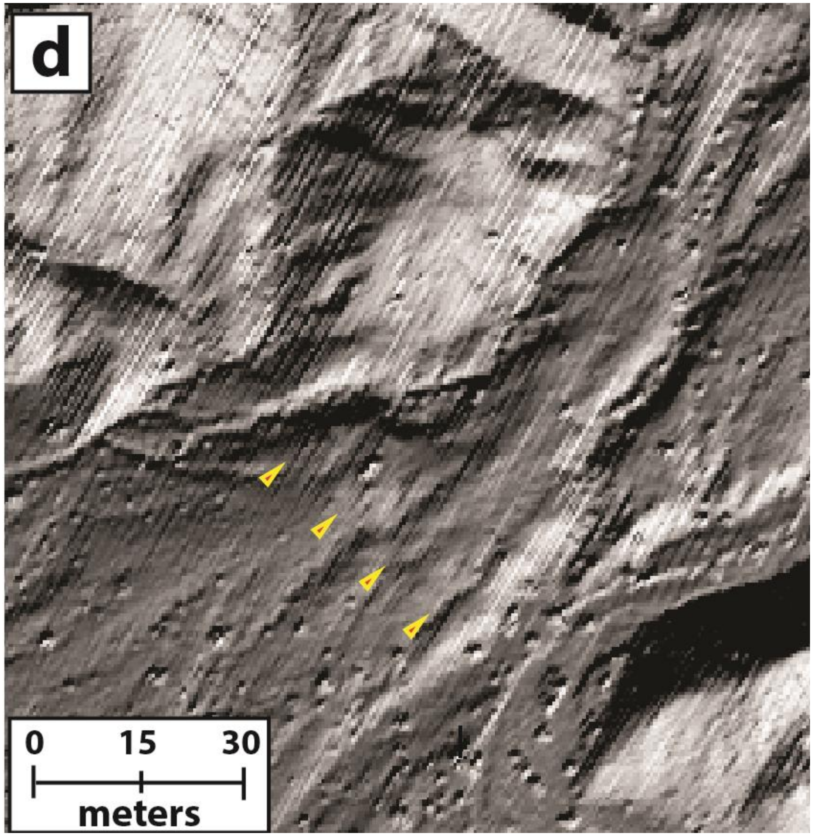


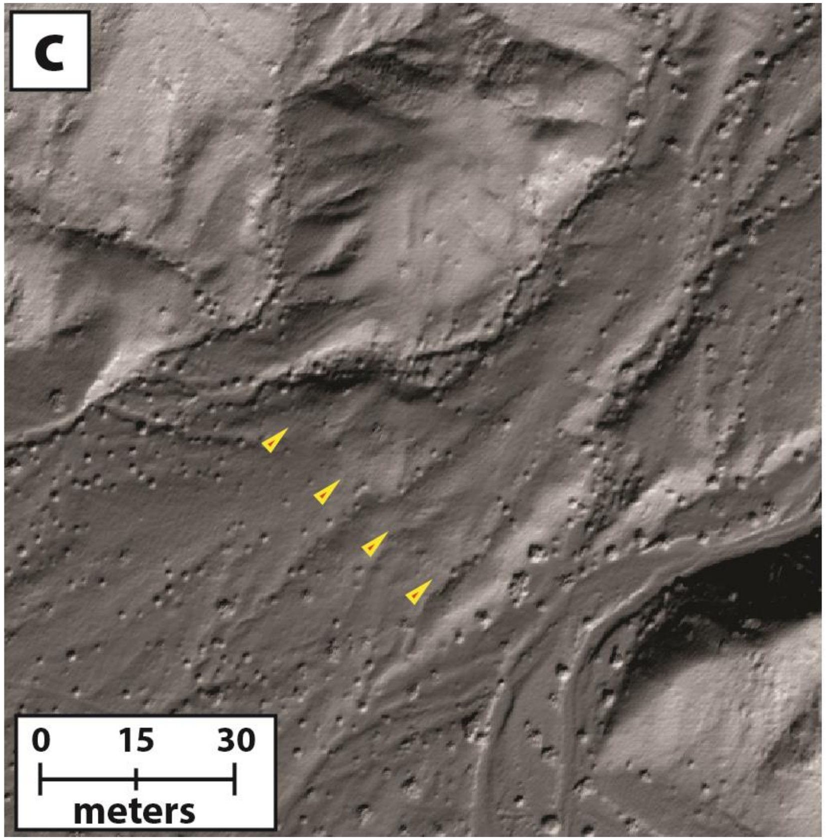
SfM ~700 pts/m^2

5 cm resolution DEM

Johnson *et al.* (2014), Rapid mapping of ultrafine fault zone topography with structure from motion, *Geosphere*

Resolution and precision of SfM topography





B4 LiDAR ~4 pts/m^2

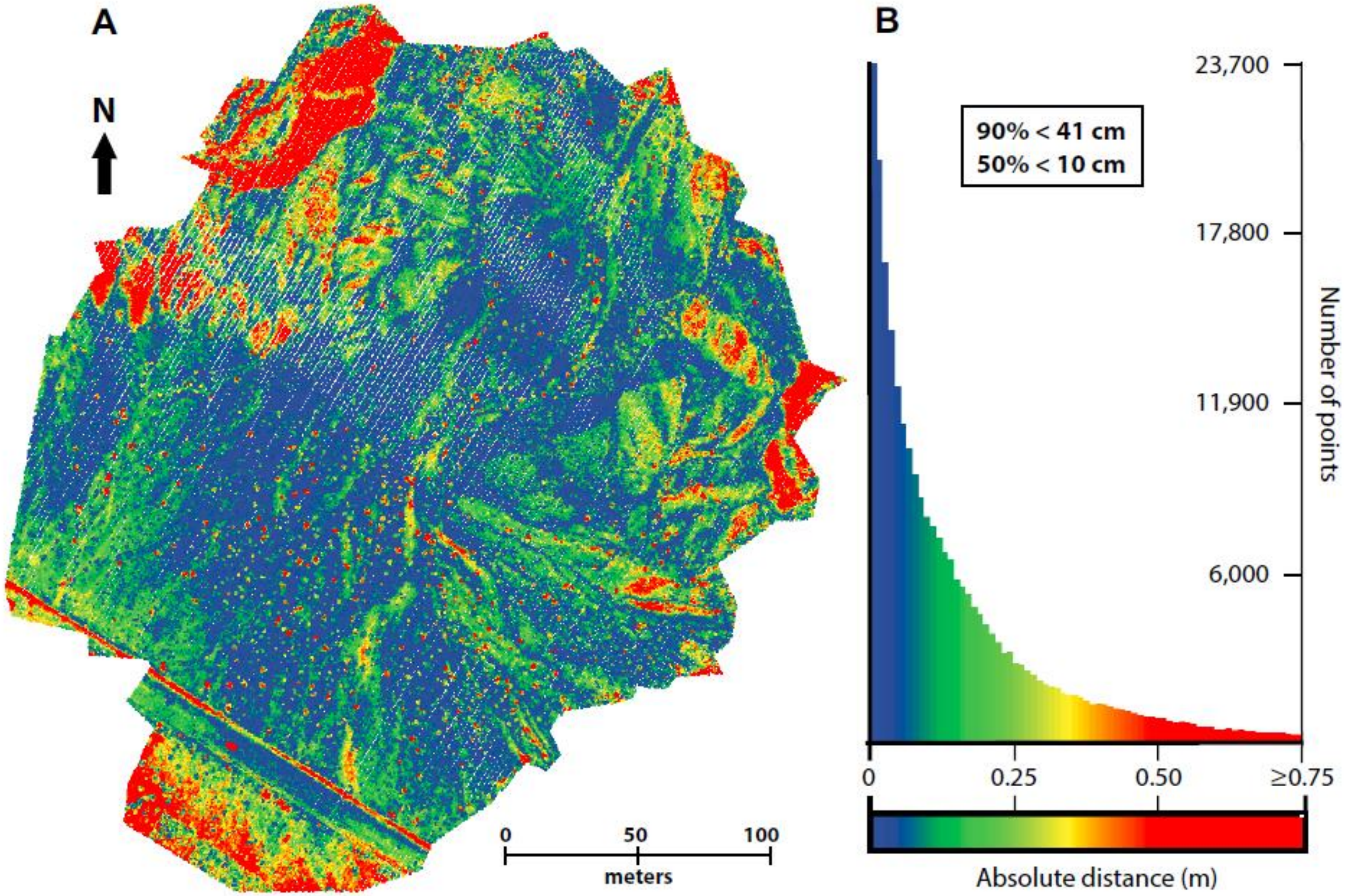
0.5 - 1 m resolution DEM

SfM ~700 pts/m^2

5 cm resolution DEM

Johnson *et al.* (2014), Rapid mapping of ultrafine fault zone topography with structure from motion, *Geosphere*

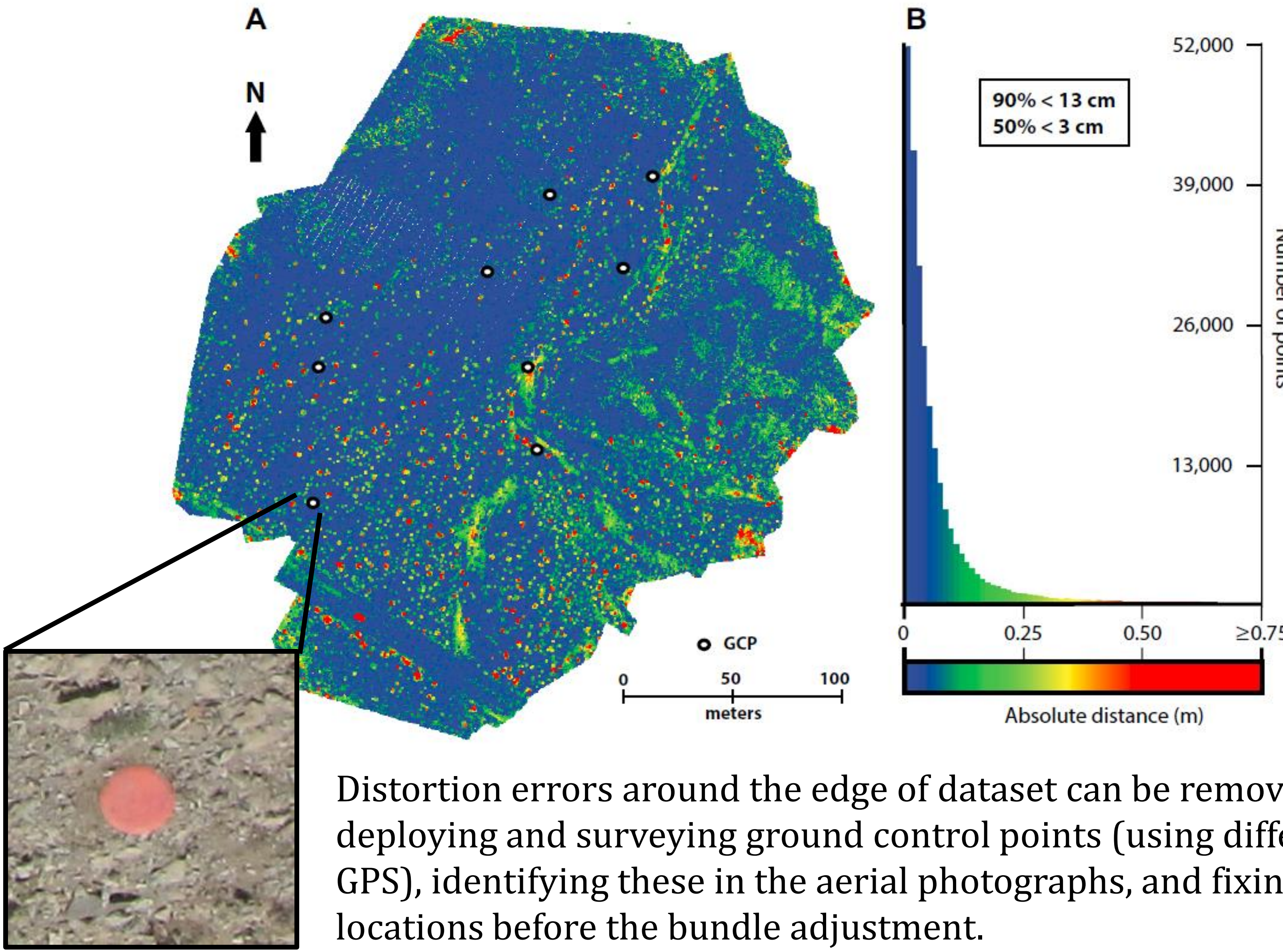
Resolution and precision of SfM topography



Note errors of >50 cm concentrated around edge of dataset. These probably reflect a trade-off in the bundle adjustment between estimates of the radial distortion of the camera lens and the topography

Johnson *et al.* (2014), Rapid mapping of ultrafine fault zone topography with structure from motion, *Geosphere*

Resolution and precision of SfM topography



Distortion errors around the edge of dataset can be removed by deploying and surveying ground control points (using differential GPS), identifying these in the aerial photographs, and fixing the locations before the bundle adjustment.

Johnson *et al.* (2014), Rapid mapping of ultrafine fault zone topography with structure from motion, *Geosphere*