Structure from Motion Summary

C. Symeonidis, I. Karakostas, Prof. Ioannis Pitas
Aristotle University of Thessaloniki
F. Zhang, D. Hall, T. Xue, S. Boyle D. Bull
University of Bristol
pitas@csd.auth.gr
www.aiia.csd.auth.gr

Version 2.4

Artificial Intelligence & Information Analysis Lab

Structure from Motion

- **Image-based 3D Shape Reconstruction**
- Structure from motion
- Structure from motion applications
- 3D Shape reconstruction workflow issues

Image-based 3D Shape Reconstruction

- A single monocular image does not convey depth information.
- But it can be used detect points at any range.

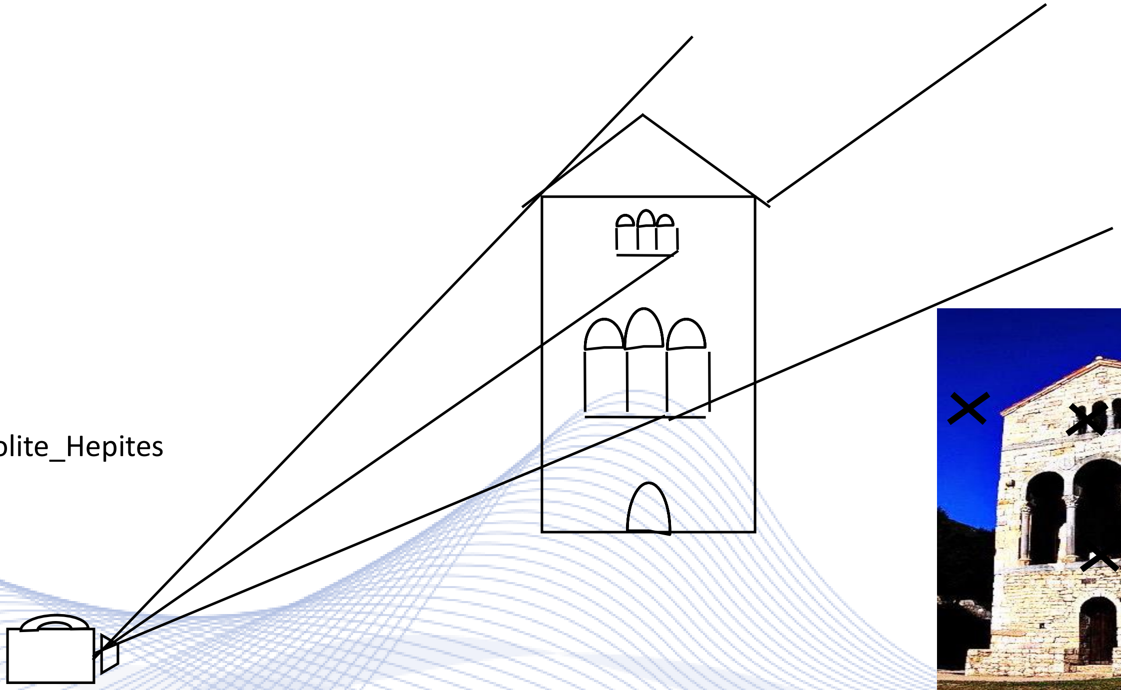
Calibrated monocular image

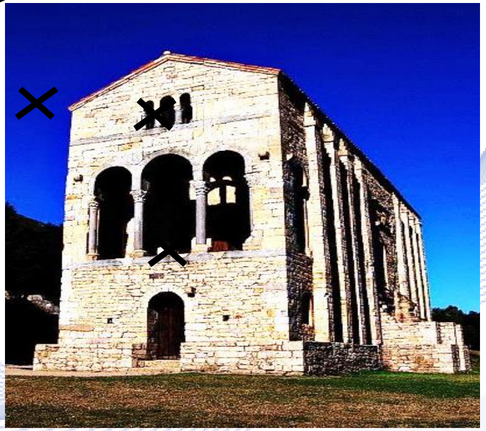
The camera detects:

- Azimuth and elevation angles per pixel, with accuracy ranging from 0.1 to 0.01 degrees.
- Colour of the reflected or emitted light by the scene point per pixel.
- Millions of pixels per image.
- Tens of images per second.

Calibrated monocular image

Victor Blacus
(https://commons.wikimedia.org/wiki/File:Amagnetic_theodolite_Hepites_1.jpg),
"Amagnetic theodolite Hepites 1",
https://creativecommons.org/licenses/by-sa/3.0/legalcode



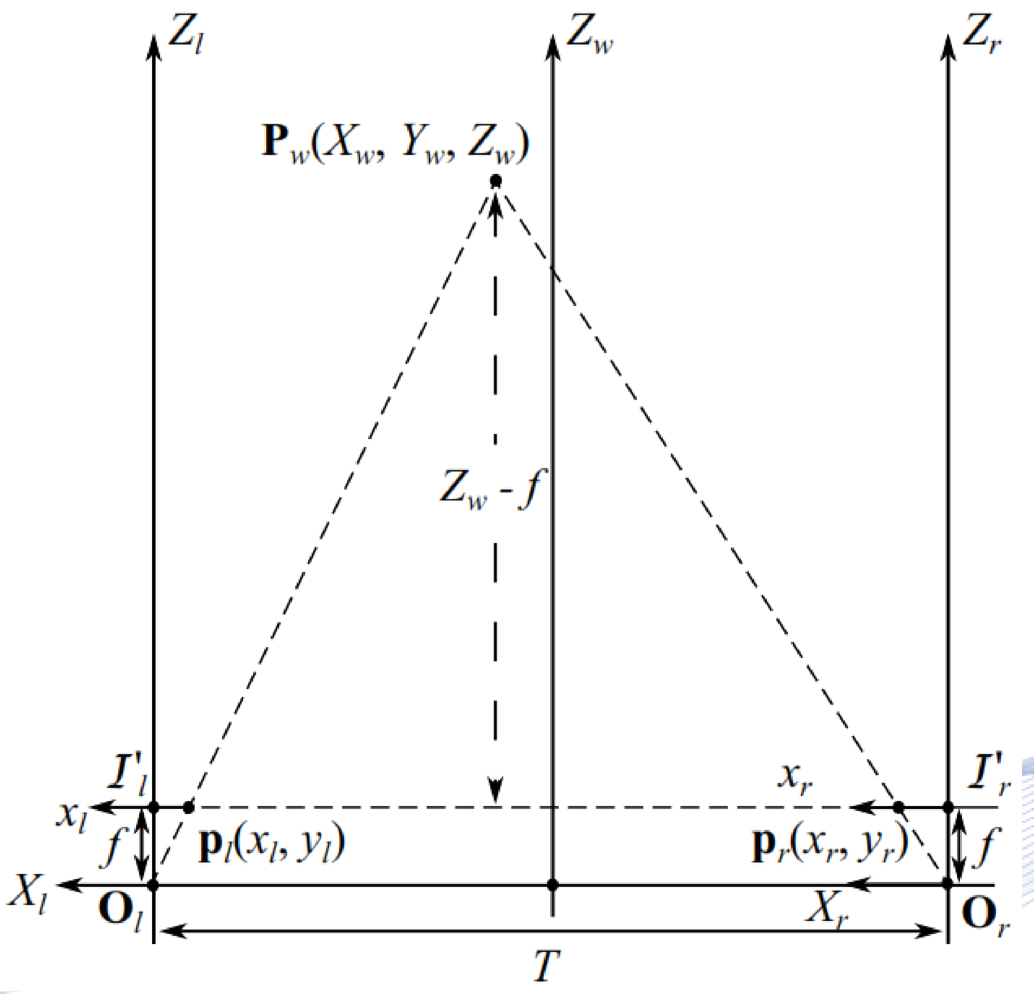


Ángel Miguel Sánchez
(https://commons.wikimedia.org/wiki/File:Sta_Maria_Naranco.jpg),
"Sta Maria Naranco", modified,
https://creativecommons.org/licenses/by-sa/3.0/es/deed.en

Basics of Stereopsis

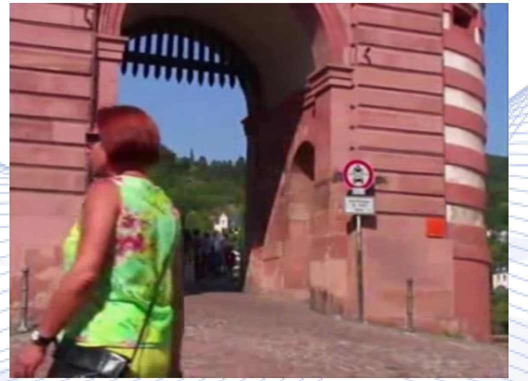
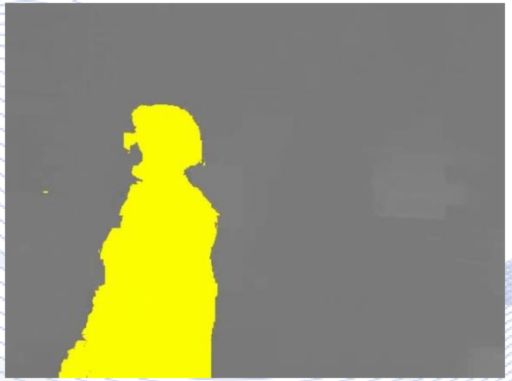
Parallel Stereo vision Geometry

T: baseline
f: focal length



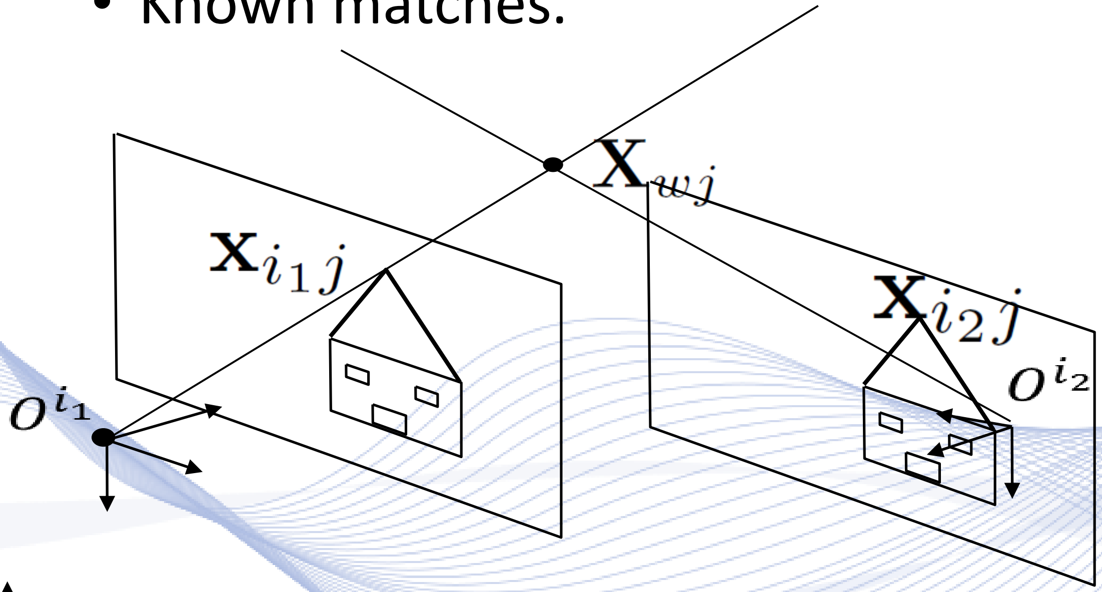
Stereo imaging

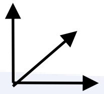
- Two cameras in known locations.
- Calibrated cameras.
- Stereo images can create a disparity (depth) map.

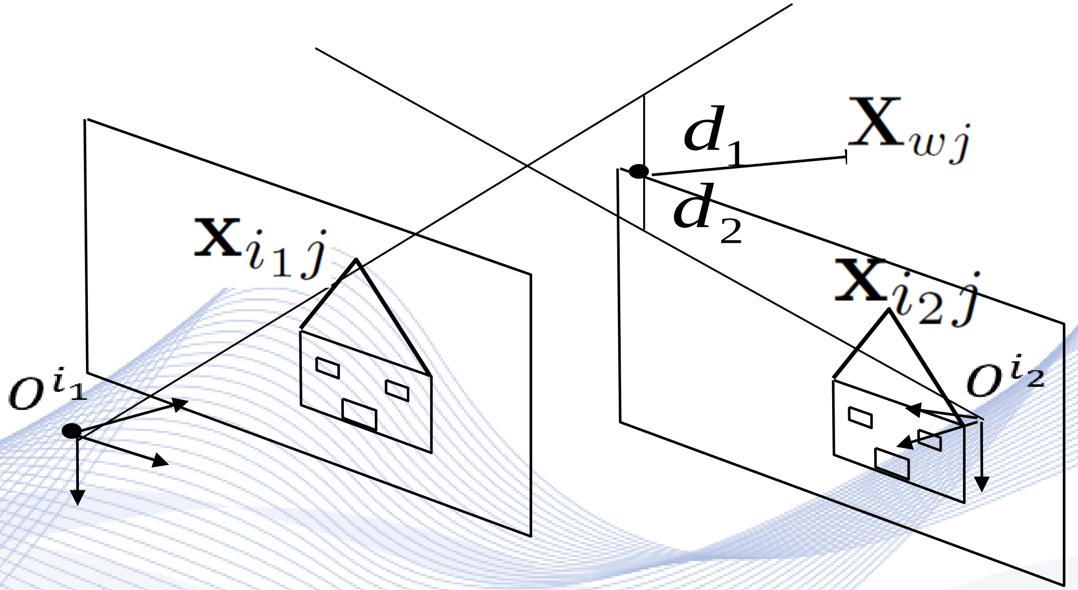
3D perception (at least two views)

- Two cameras in known locations.
- Calibrated cameras.
- Known matches.





In an ideal world ...



In this real world ...

Structure from Motion

- Image-based 3D Shape Reconstruction
- **Structure from motion**
- Structure from motion applications
- 3D Shape reconstruction workflow issues

Structure from Motion (SfM)

- Unknown camera location/orientation.
- Cameras can be fully, partially or non-calibrated.
- Unknown feature correspondences across views.
- Computation up to scale factor:
  - Camera location.
  - 3D location of the matched feature points.

Structure from Motion (SfM)

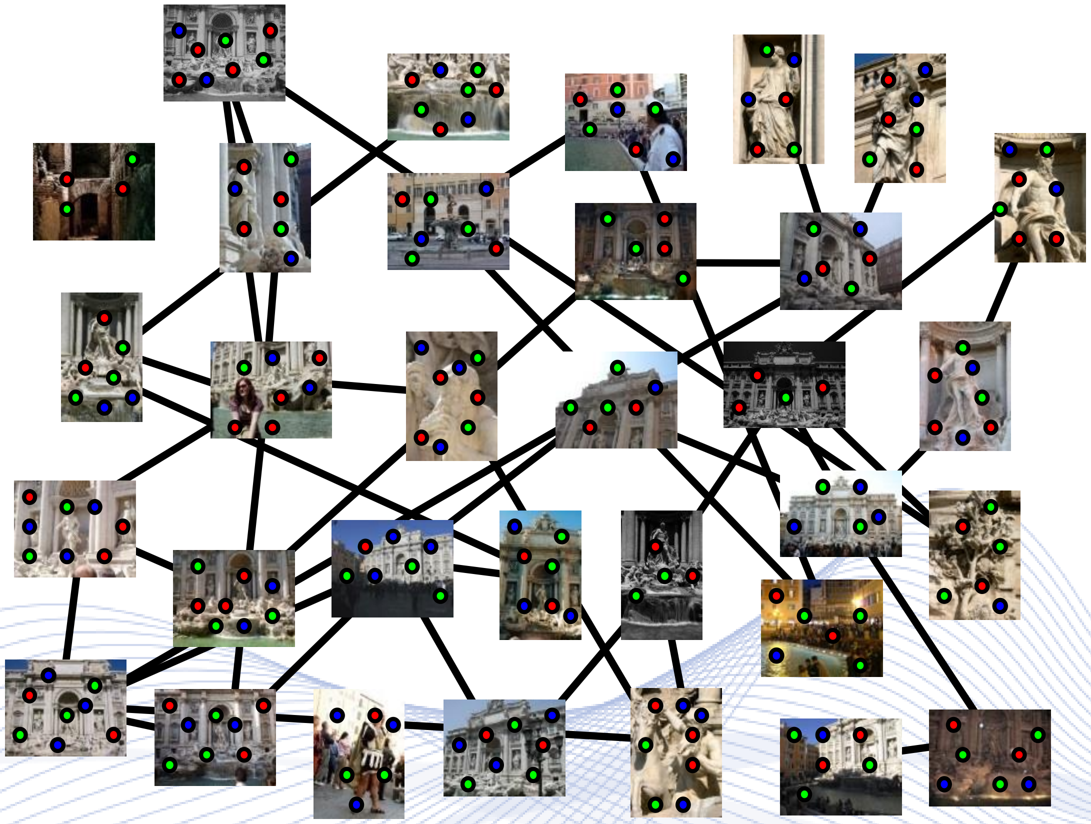


Photo tourism: exploring photo collections in 3D (https://www.youtube.com/watch?v=6eQ-CB8TY2Q)

N Snavely, SM Seitz, R Szeliski. "*Modeling the world from internet photo collections",* International Journal of Computer Vision, 80 (2), 189-210
Hartley, Richard, and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.

Structure from Motion (SfM)

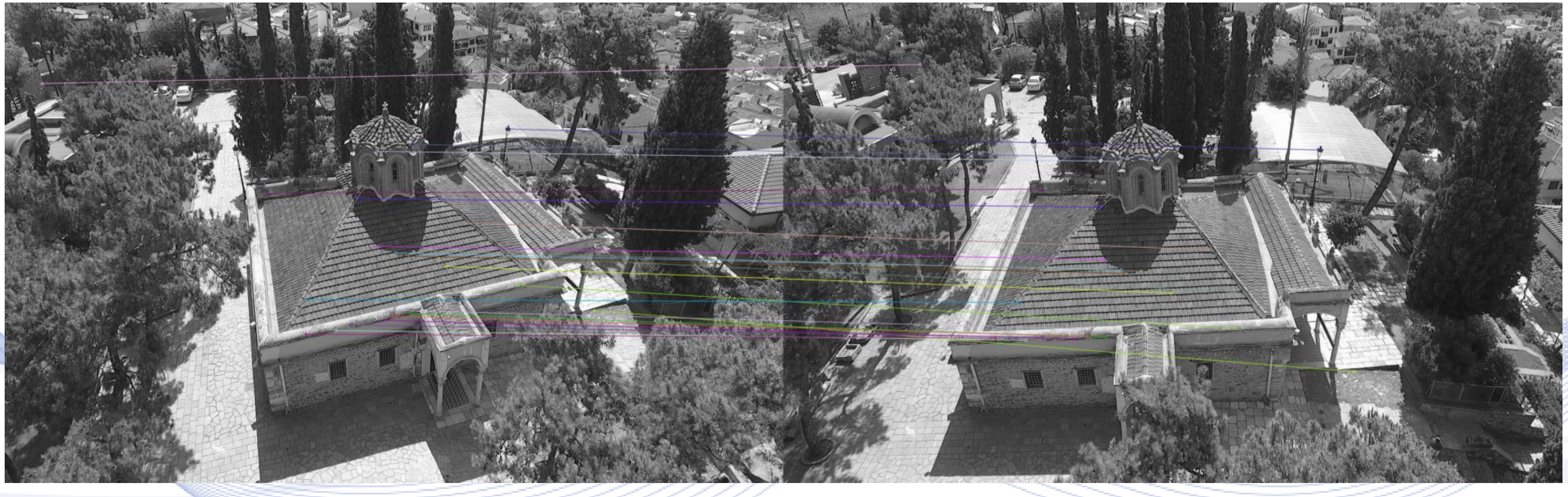
- The three-dimensional (3D) scene structure from a set of camera images is known in the computer vision community as Structure from Motion (SfM).
- Some basic steps of SfM are:
  - feature extraction
  - feature matching
  - triangulation and bundle adjustments

Structure from Motion (SfM)

- Structure from Motion (SfM) perfoms two tasks simultaneously:
    - 3D scene geometry reconstruction from a set of camera images and
    - Camera calibration.
- Images can be cquired by:
    - multiple synchronized cameras or
- one moving camera, or unsyncroniz3ed multiple cameras, if the scene and illumination are static.

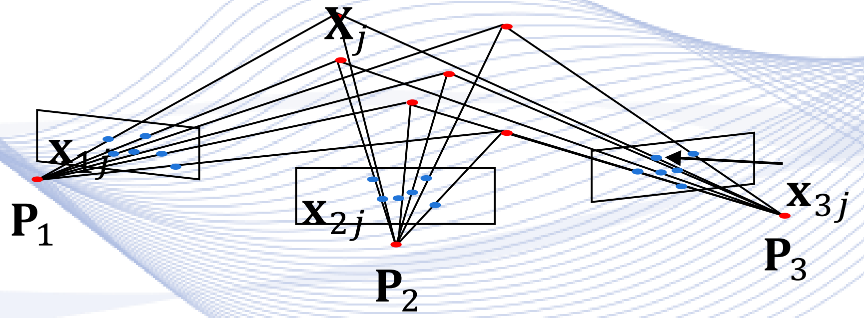
Feature extraction and matching (SfM)

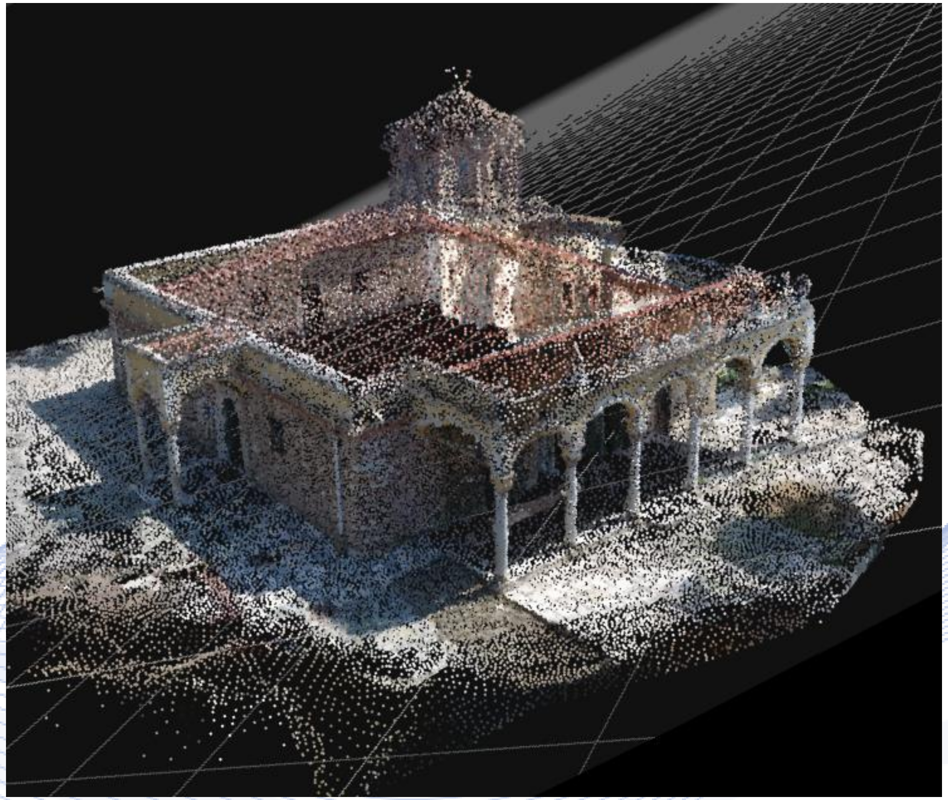
- Extract keypoints based on local features.
- Common feature extractors are SIFT, SURF, ORB etc.
- The keypoints are matched between images taken from different views.



Triangulation and Bundle Adjustment (SfM)

- Bundle adjustment and triangulation are the final steps to estimate the camera parameters and create an accurate point cloud.
- Further techniques are used to make the point cloud denser and to reconstruct surface based on it.





Structure from Motion (SfM)

- Initial SfM stages end up providing an accurate initial guess to non-linear re-projection error optimization (Bundle Adjustment):

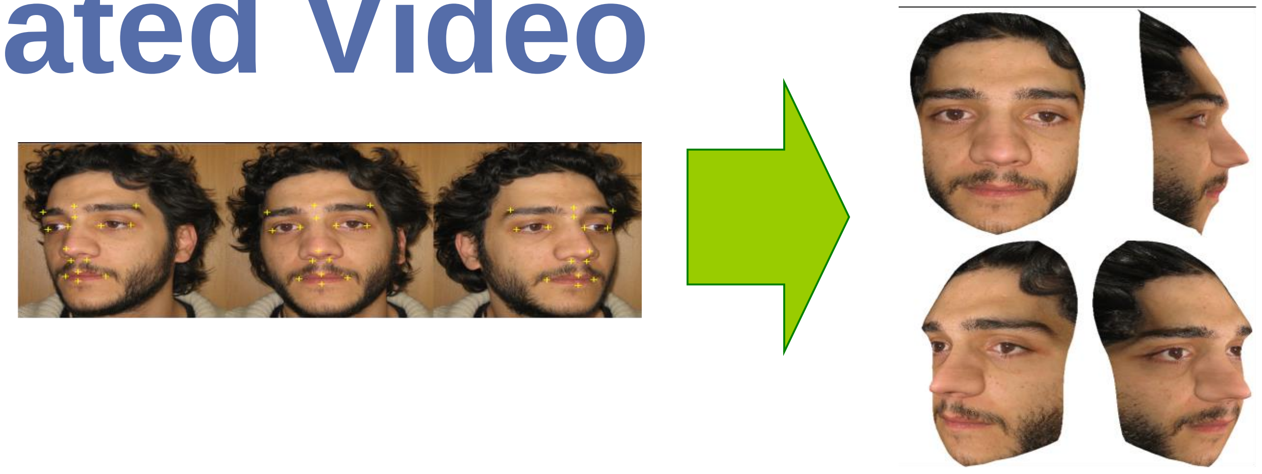
$$\underset{\mathbf{T}_{iw},\mathbf{X}_{wj}}{\operatorname{argmin}} \sum_{i,j} \rho \left(\left\| \mathbf{x}_{ij} - \pi_i \left(\mathbf{T}_{iw}, \mathbf{X}_{wj} \right) \right\|^2 \right)$$

Structure from Motion

- Image-based 3D Shape Reconstruction
- Structure from motion
- **Structure from motion applications**
- 3D Shape reconstruction workflow issues

SfM in 3D Face Reconstruction from Uncalibrated Video

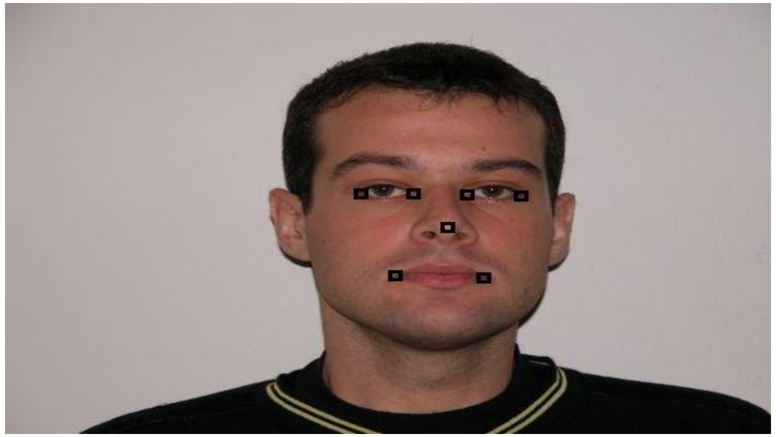
Problem statement:

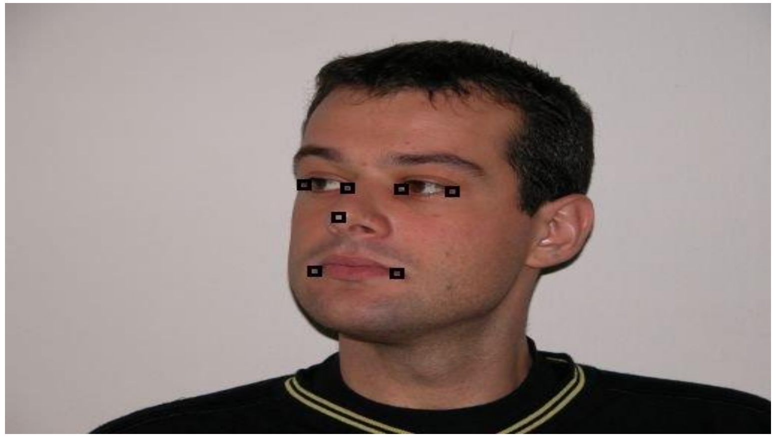


- Input: facial images or facial video frames, taken from different view angles, provided that the face neither changes expression nor speaks.
- Output: 3D face model (saved as a VRML file) and its calibration in relation to each camera. Facial pose estimation.
- Applications:

3D face reconstruction, facial pose estimation, face recognition, face verification.

SfM in 3D Face Reconstruction

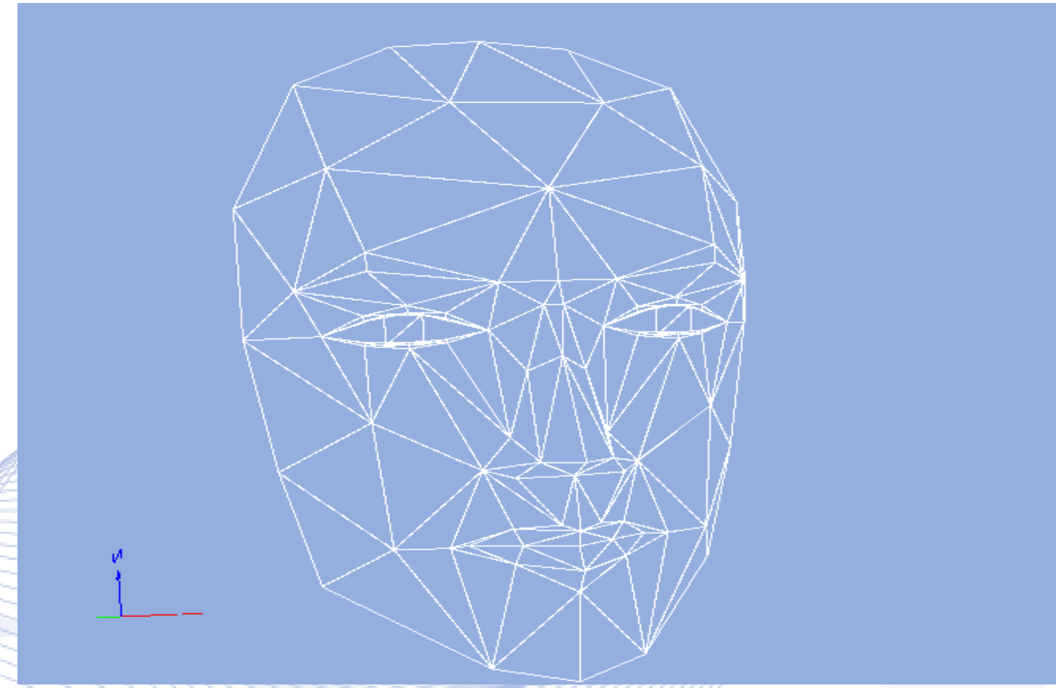




Input: three images with a number of matched characteristic feature points.

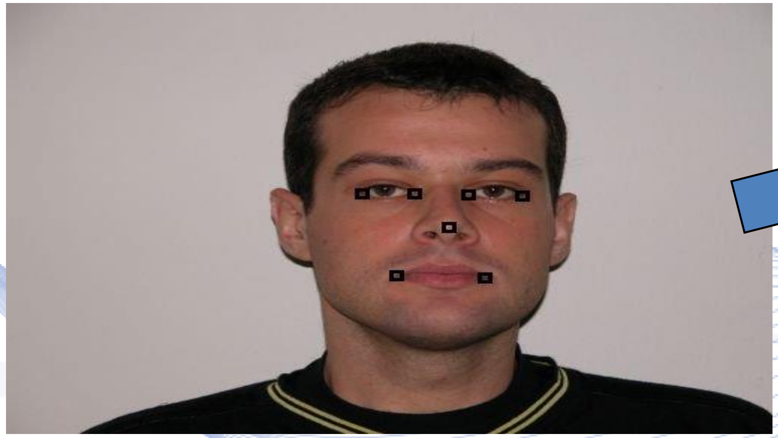
3D Face Reconstruction

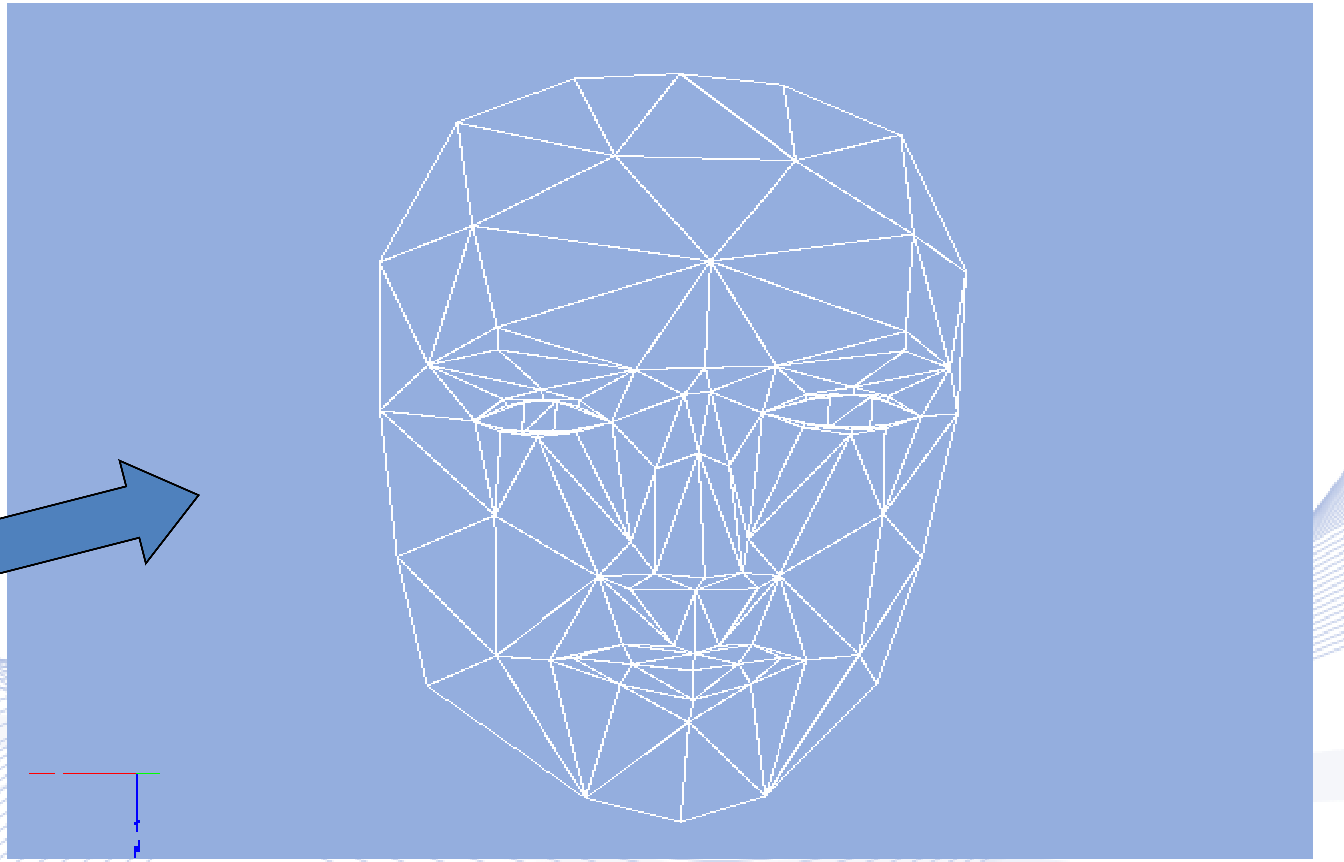
- The CANDIDE face model has 104 nodes and 184 triangles.
- Its nodes correspond to characteristic points of the human face, e.g. nose tip, outline of the eyes, outline of the mouth etc.



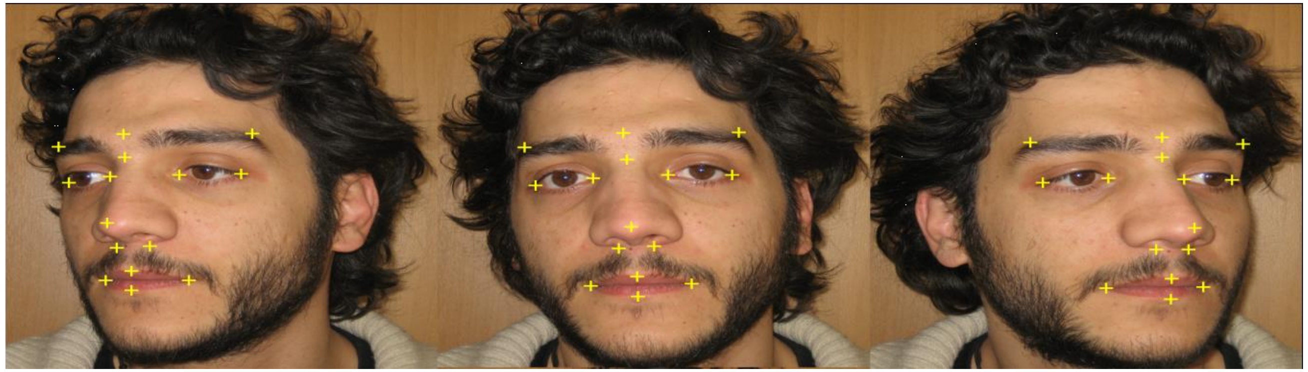
SfM in 3D Face Reconstruction

Example





SfM in 3D Face Reconstruction



Selected features

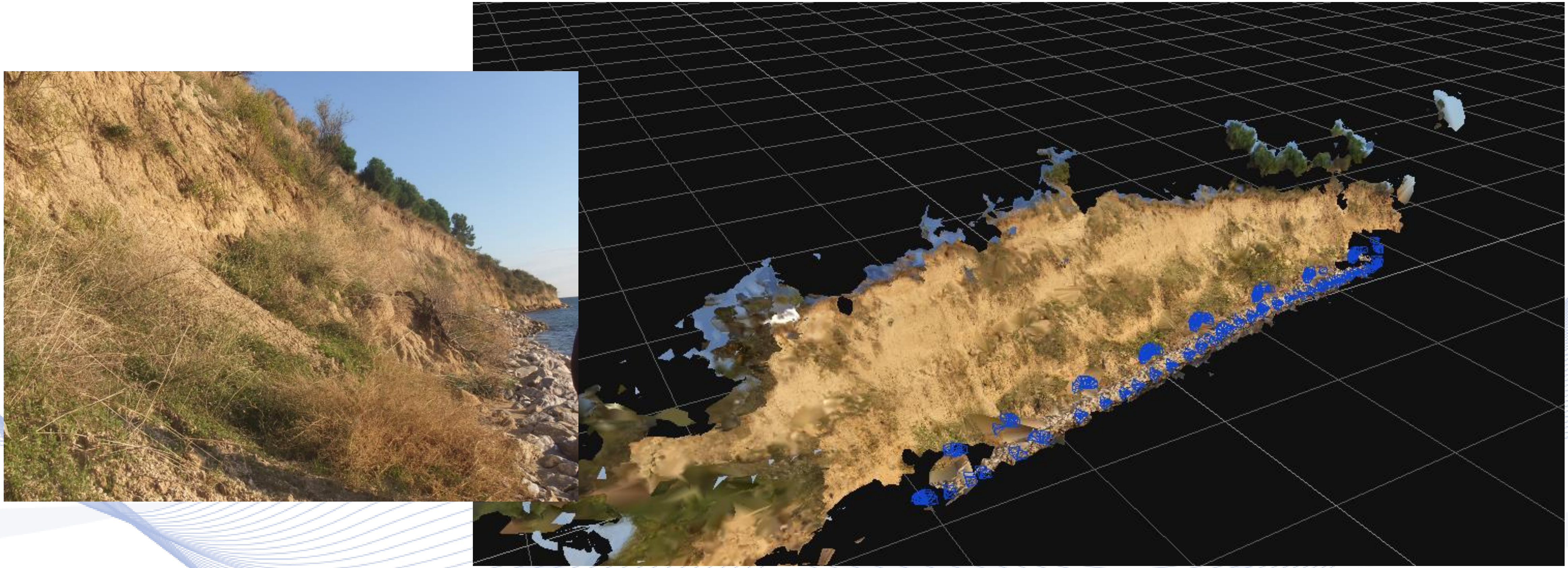
CANDIDE grid reprojection

3D face reconstruction

SfM in 3D landscape reconstruction

- Cliff images

SfM in 3D landscape reconstruction



- 3D Cliff surface reconstruction

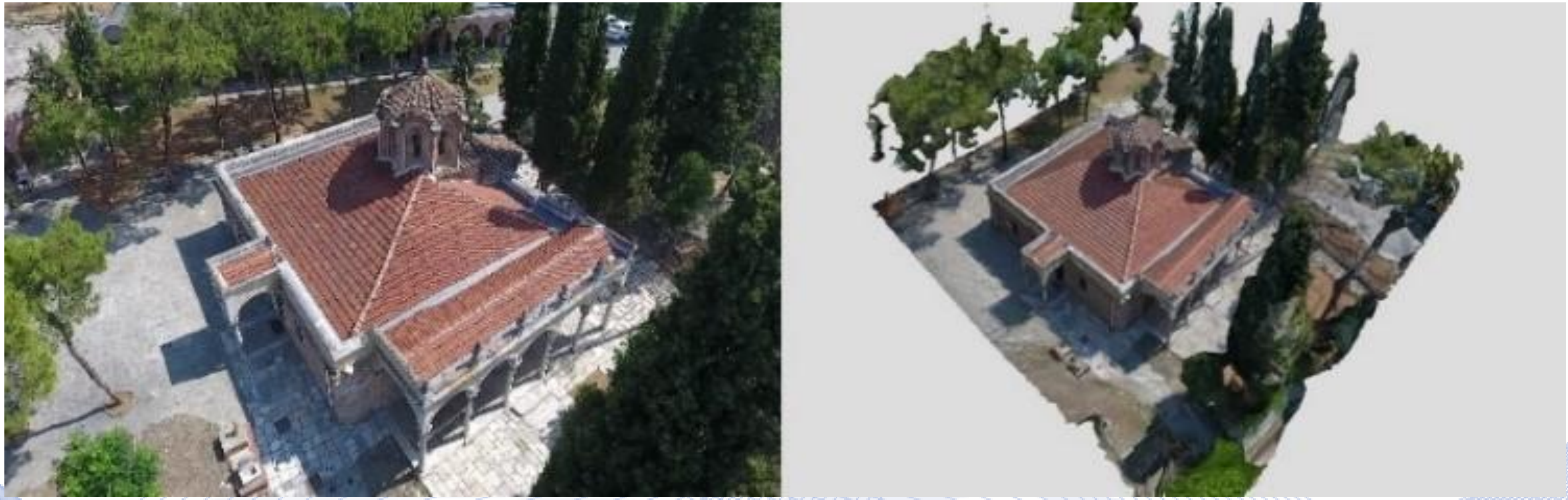
3D reconstruction using images from drones

- Two of the fifteen Paleochristian and Byzantine monuments of Thessaloniki that were included in the UNESCO World Heritage List:
  - The Vlatadon Monastery
  - The Church of Saint Nicholas Orphanos
- Imaging by mostly orbiting a drone around them at different heights, respecting always the corresponding flight regulations:
  - avoiding collisions with nearby objects (trees, buildings, etc.) had a negative impact in capturing the close-by details of each monument.
  - Additional image collection from ground cameras.
- 3DF Zephyr: a commercial software for SoA 3D scene reconstruction.

SfM in 3D building reconstruction

- Vladaton monastery

SfM in 3D building reconstruction



SfM in 3D building reconstruction

Copyright Hellenic Ministry of Culture and Sports (L. 3028/2002)

Structure from Motion

- Image-based 3D Shape Reconstruction
- Structure from motion
- Structure from motion applications
- **3D Shape reconstruction workflow issues**

Image Capturing from UAVs

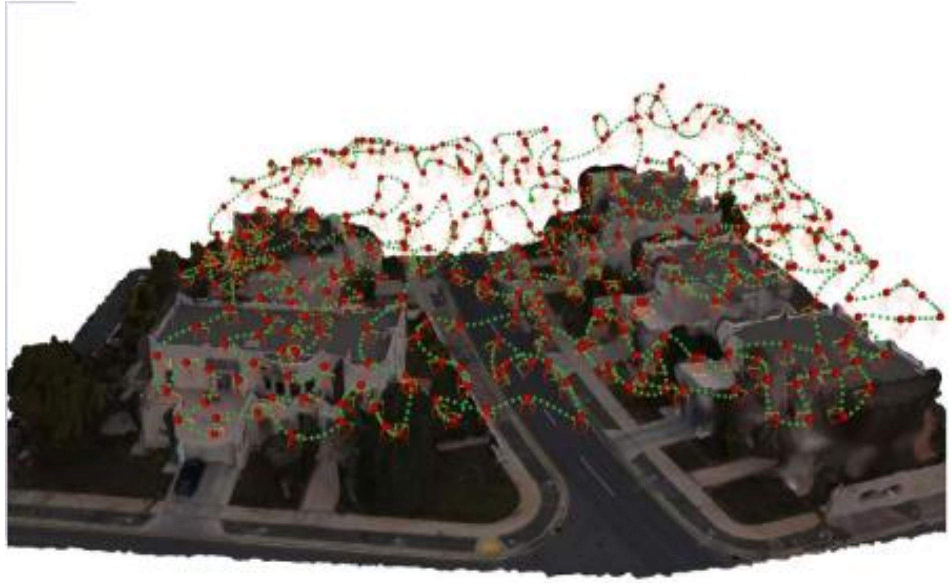
Motivation

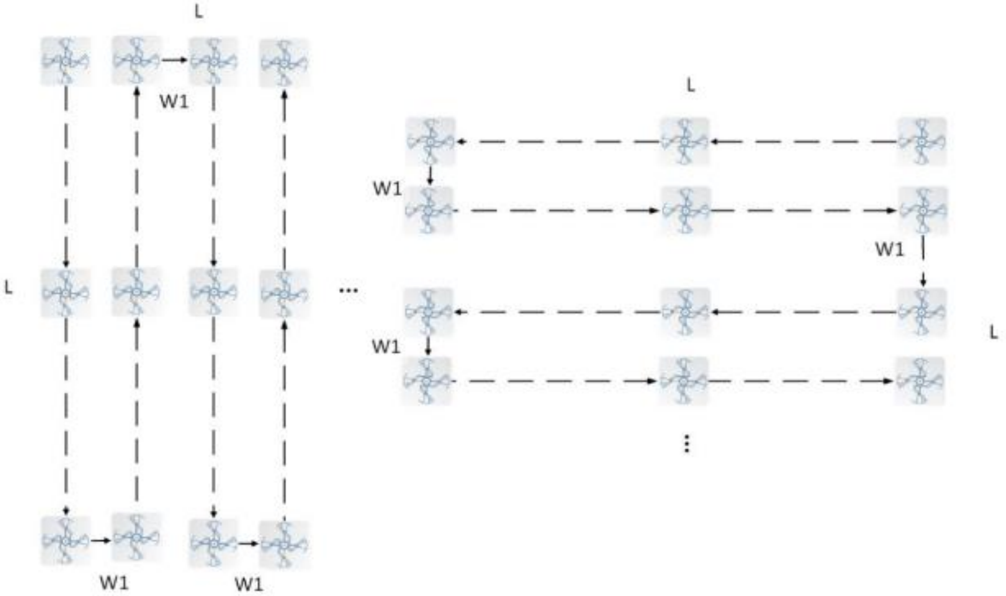
- Multiple view-point photo scanning provides the basis (2D data) in photogrammetry for 3D model reconstruction.
- Environmental images can be captured using real drones or obtained by taking screen shots within software such as Google Earth.
- The target of this step is to optimize the scanning strategy to obtain as few images as possible to achieve the optimal reconstruction quality.

Scanning Strategy

- Information required:
  - Location of the environmental area.
  - Camera parameters (sensor size and focal length).
- Optimal shot parameters:
  - Flight trajectory
  - Flying heights
  - Viewing angles
  - Image overlap ratios (the number of images).

Optimal UAV Flight Trajectory





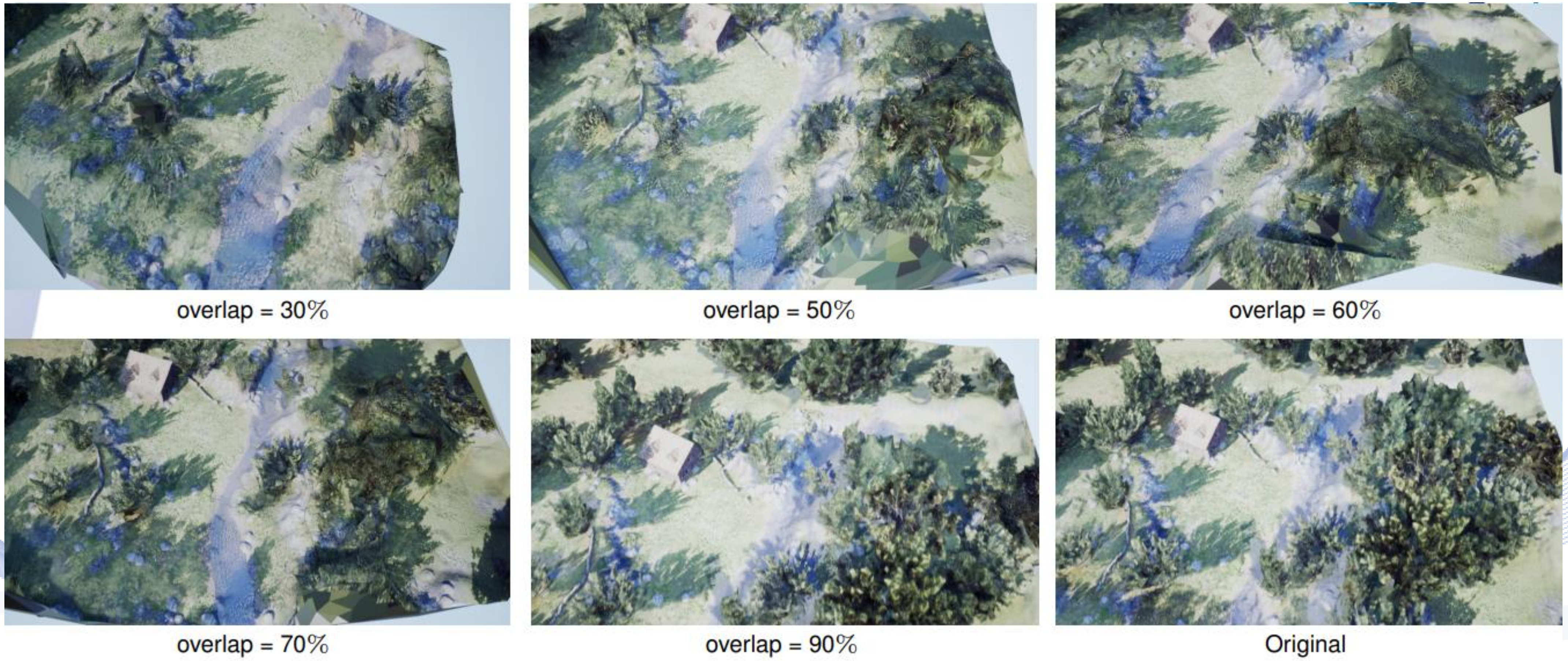
- The optimal flight trajectory for a specific background environment is **highly dependent on the given landscape and object complexity** [Smith et al., 2018]
- The most commonly used flight pattern in practice is **grid scanning** in two orthogonal horizontal directions.
- We have employed a grid scanning strategy in order to simplify the scanning strategy for both shooting with real drones and capturing within virtual globe software packages

UAV Camera Viewing Angles

- Viewing angle is also denoted as **gimbal rotation angle** in drone cinematography.
- The reconstruction results for three sets (for three height levels) of different gimbal rotation angles were compared for comparison, including 90/67.5/45, 85/60/35, and 70/47.5/25 degree.
- Other parameters such as flight heights and the number of images are fixed.
- The Countryside asset from UE4 market place were employed as source and also as ground truth for benchmarking.
- The captured images for each test set have been employed as inputs to generate 3D environmental models.
- Based on the subjective results, we recommend to use the intermediate angle set (85/60/35 degree) in practice.

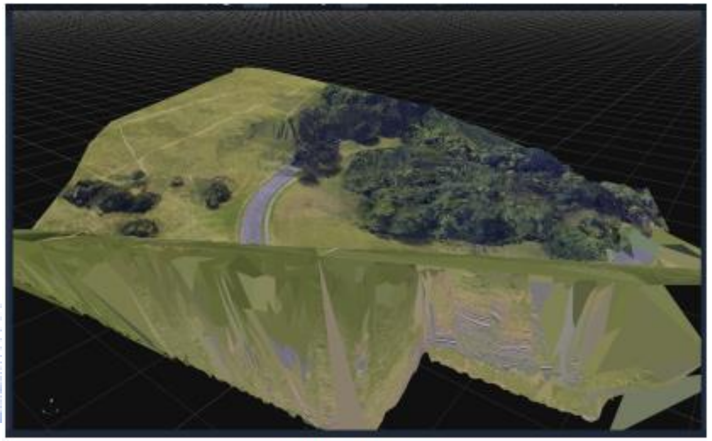
(a) High angles (b) Low angles (c) Intermediate angles

Image Overlap Ratio

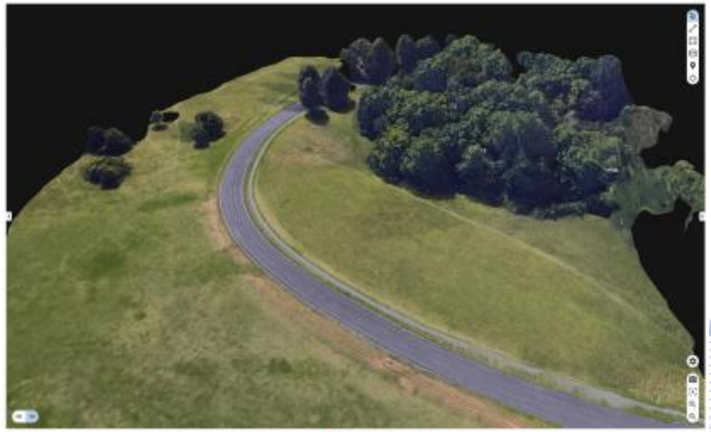


Photogrammetry Reconstruction

- Various 3D photogrammetry software packages can be used to produce 3D models.
- AutoDesk produces poor results from Google Earth captured photos and requires ‘cloud credits’ to perform analyses remotely on Autodesk servers.
- 3DF Zephyr Aerial can generate 3D models with improved quality but was mainly designed for object-based scenarios.
- Pix4D Mapper was developed for background environment survey and with improved reconstruction performance for landscapes and lower computation complexity (50% of 3DF Zephyr).



3DF Zephyr



Pix4D

Pre/post-processing for 3D Reconstruction

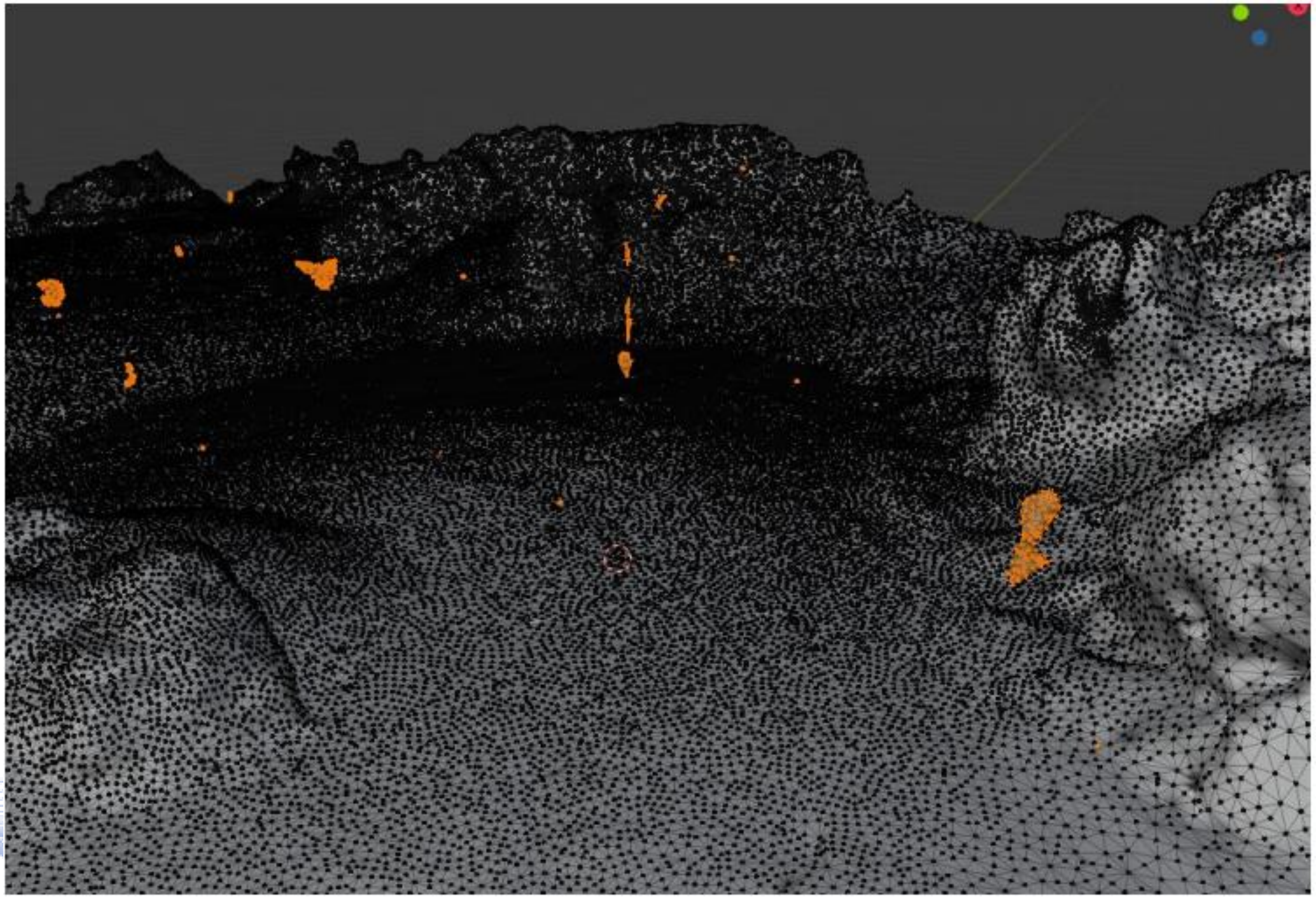
Pre-processing

- The input image dataset may contain artefacts due to photogrammetry errors or moving objects, which can result in significant distortions during reconstruction.
- These defect images can be removed or corrected through texture in-painting [Tschumperle and Deriche, 2005; Wong and Orchard, 2008].
- Most of existing in-painting algorithms are relatively complex and will be extremely time consuming to process a large number of images.
- Simple manual outlier rejection was applied in this work on the input image dataset in order to achieve efficient reconstruction.

Post-processing

- After pre-processing, photogrammetry software can provide reasonably good results, but there are still a large amount of noticeable artefacts, e.g. bumps and holes, which could affect viewing experience.
- 3D model editing can be employed to further correct these distortions.
- The following features have been used to enhance the 3D texture mesh:
    - Isolated mesh component removal and flat surface smoothing.
    - Mesh modification and texture paint.

Isolated Mesh Component Removal



Flat Surface Smoothing

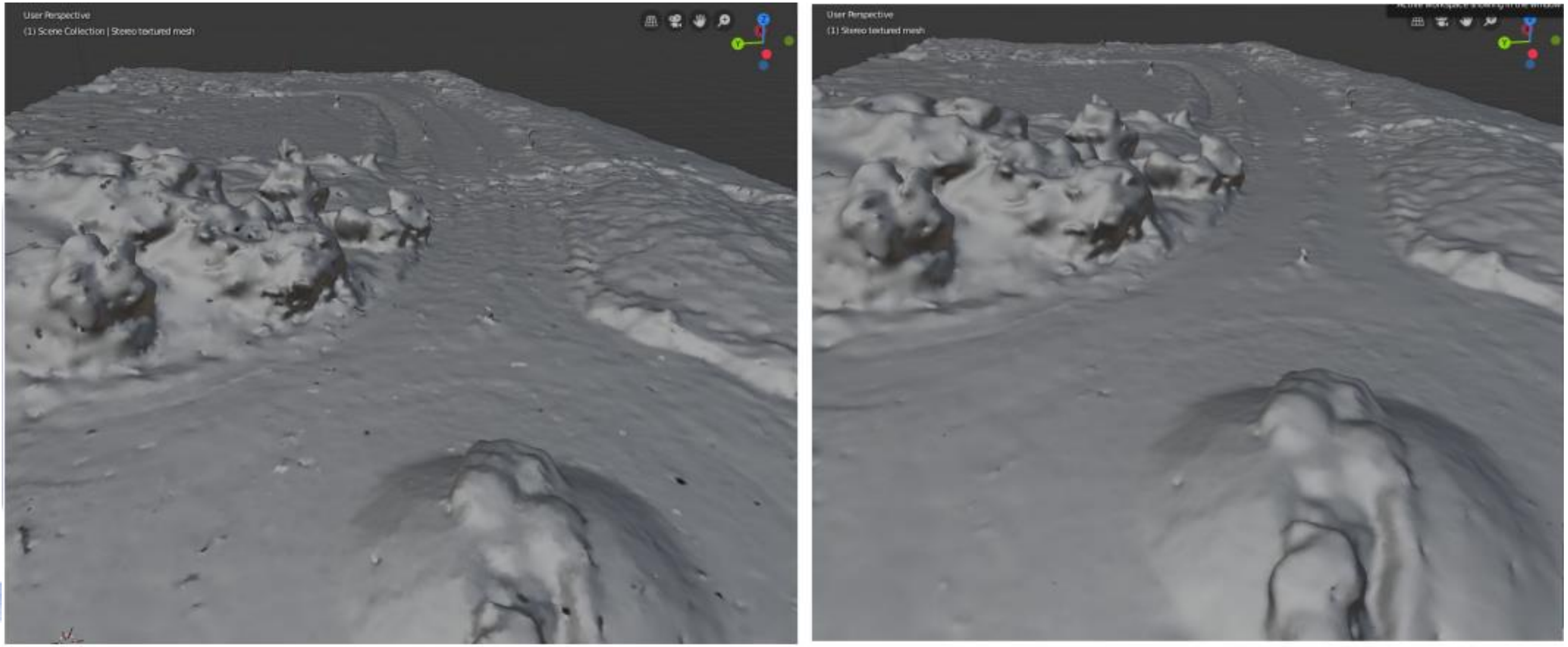


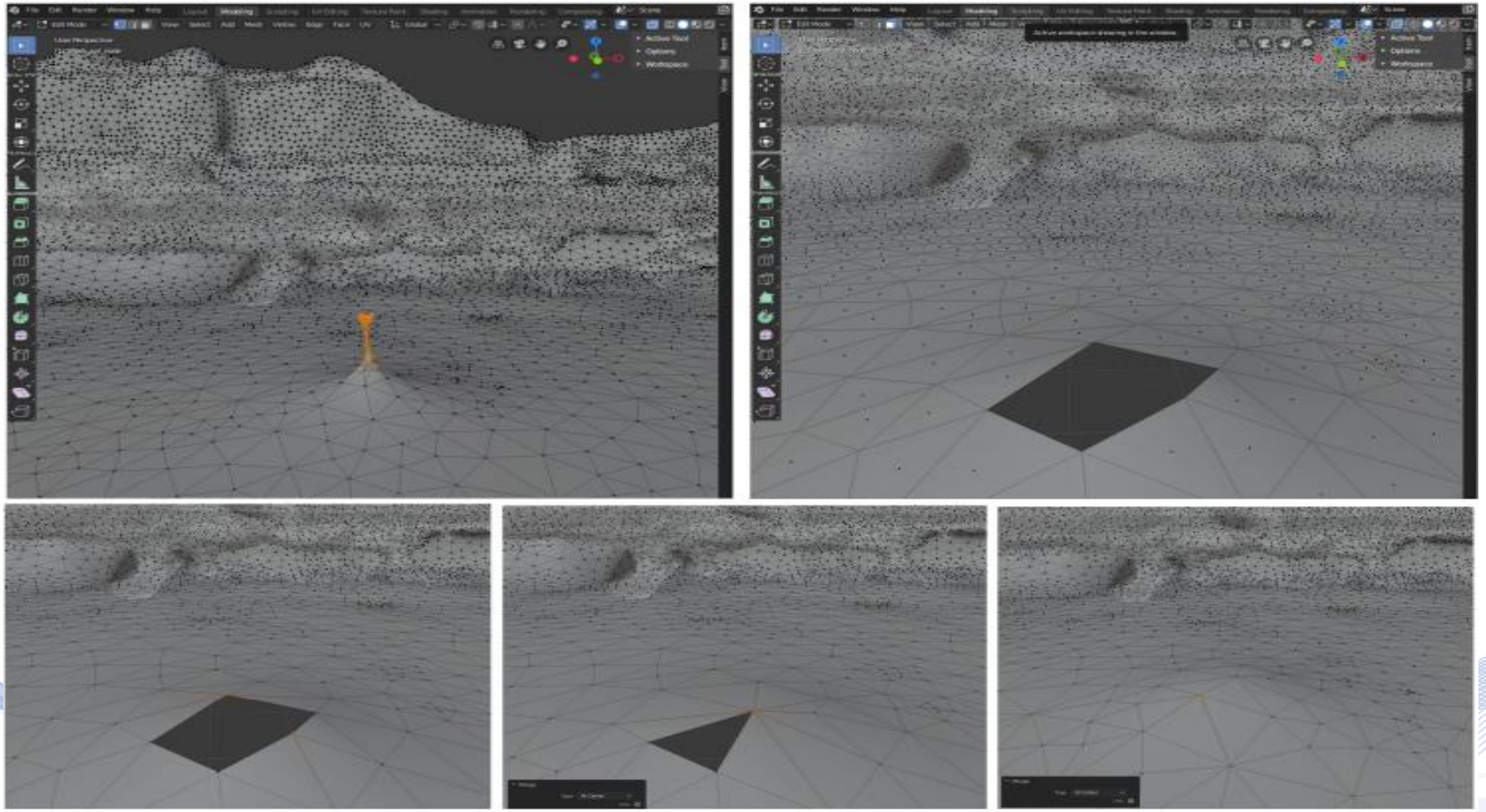
Figure: An example of Smoothing. (Left) The initial 3D model structure. (Right) The processed 3D model structure after Smoothing.

Artificial Intelligence & Information Analysis Lab

University of BRISTOL

VILab Visual Information Laboratory

Mesh Modification



Texture Paint

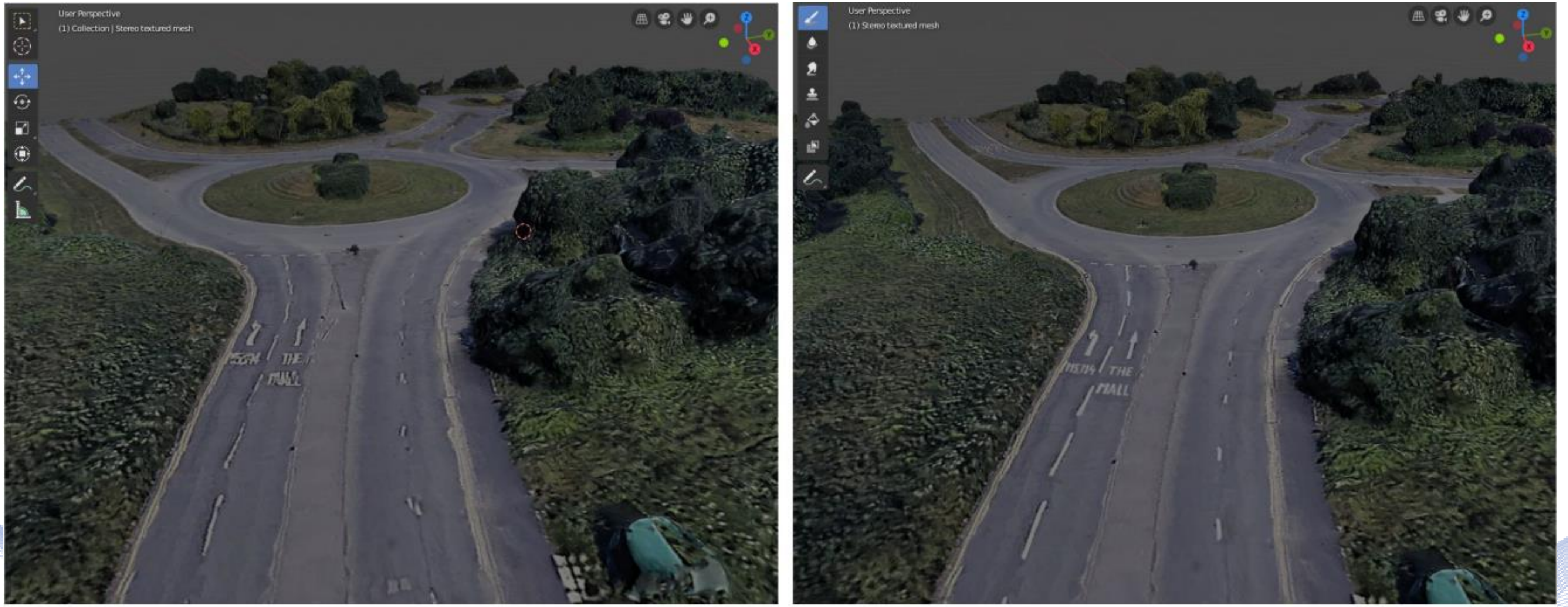


Figure: (Left) A screen shot of the 3D Model before applying Texture Paint. (Right) A screen shot of the 3D Model after applying Texture Paint.

Bibliography

[1] L. Arbace, E. Sonnino, M. Callieri, M. Dellepiane, M. Fabbri, A. I. Idelson, R. Scopigno, "Innovative Uses of 3D Digital Technologies to Assist the Restoration of a Fragmented Terracotta Statue", Journal of Cultural Heritage, 14(4), 332-345, 2013.

[2] H. Bay, A. Ess, T. Tuytelaars, L. Gool, "Speeded-Up Robust Features (SURF)", Computer Vision and Image Understanding, 110(3), 346-359, 2008.

[3] J. D. Boissonnat, "Geometric Structures for Three-dimensional Shape Representation", ACM Transactions on Graphics (TOG), pp. 266-286, 1984.

[4] M. Callieri, P. Cignoni, M. Dellepiane, R. Scopigno, "Pushing Time-of-Flight Scanners to the Limit", In Proceedings of the 10th International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage (VAST), (pp. 85-92), 2009.

[5] D. Caspi, N. Kiryati, J. Shamir, "Range Imaging with Adaptive Color Structured Light", IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(5), 470-480, 1998.

[6] C. Engels, H. Stewenius, D. Nister, "Bundle adjustment rules", Bonn, Germany: Springer, 2006.

Bibliography

[7] Y. Furukawa, C. Hernández, “Multi-View Stereo: A Tutorial”, Foundations and Trends® in Computer Graphics and Vision, 9(12), 1-148, 2015.

[8] G. D. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, International Journal of Computer Vision, 60(2), 91-110, 2004.

[9] J. Salvi, J. Pagès, J. Batlle ,”Pattern Codification Strategies in Structured Light Systems”, Pattern Recognition, 37(4), 827-849, 2004.

[10] B. Triggs, B. McLauchlan, R. I. Hartley, A.W. Fitzgibbon, “Bundle Adjustment — A Modern Synthesis”, Berlin, Germany: Springer-Verlag, 2020.

[11] A. Vrubel, O. Bellon, L. Silva, “A 3D Reconstruction Pipeline for Digital Preservation of Natural and Cultural Assets”, In Proceedings of Computer Vision and Pattern Recognition (CVPR), (pp. 2687–2694), 2009.

[12] N. Smith, N. Moehrle, M. Goesele, W. Heidrich, “Aerial path planning for urban scene reconstruction: A continuous optimization method and benchmark”, SIGGRAPH Asia 2018 Technical Papers, page 183, 2018.

Bibliography

[13] E. Seifert, S. Seifert, H. Vogt, D. Drew, J. V. Aardt, A. Kunneke, T. Seifert, “Influence of drone altitude, image overlap, and optical sensor resolution on multi-view reconstruction of forest images”, Remote Sensing, 11(10):1252, 2019.

[14] S. Hawkins, “Using a drone and photogrammetry software to create orthomosaic images and 3d models of aircraft accident sites”, ISASI Seminar, pages 1–26, 2016.

[15] David Tschumperle and Rachid Deriche, “Vector-valued image regularization with pdes: A common framework for different applications”, IEEE transactions on pattern analysis and machine intelligence, 27(4):506–517, 2005.

[16] A. Wong and J. Orchard, “A nonlocal-means approach to exemplar-based inpainting”, IEEE International Conference on Image Processing, pages 2600–2603, 2006.

Q & A

Thank you very much for your attention!

Contact: Prof. I. Pitas
pitas@csd.auth.gr