Structure from Motion

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 3.4.1

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

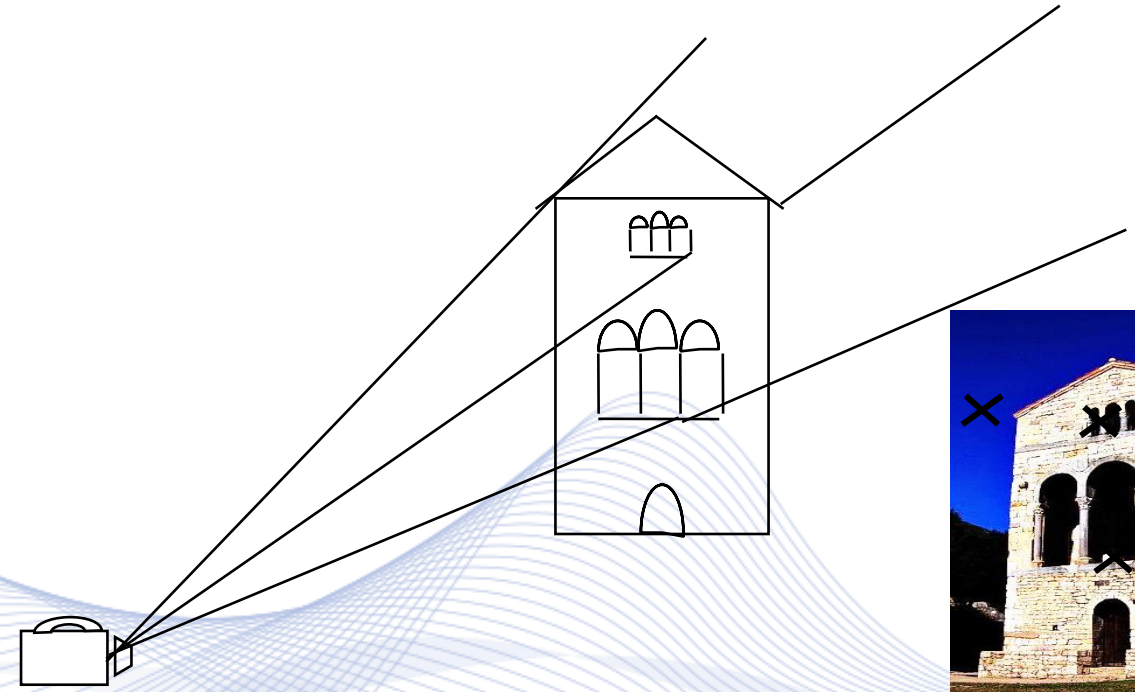
The camera detects:

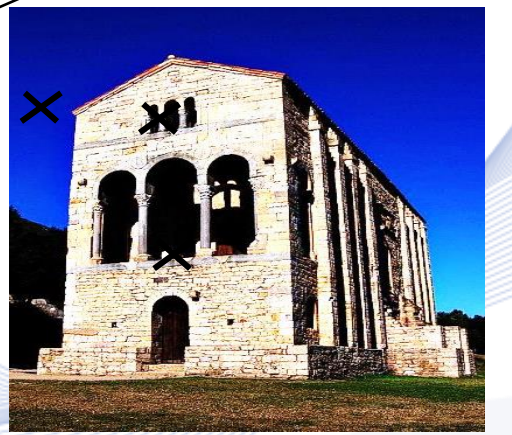
- Azimuth and elevation angles per pixel, with accuracy ranging from 0.1 to 0.01 degrees.
- Colour of the reflected or emitted light by the scene point per pixel.
- Millions of pixels per image.
- Tens of images per second.

Calibrated monocular image

Theodolite can measure both horizontal and vertical angles [BLA].

Calibrated monocular image





Ray casting [SAN].

Stereo imaging

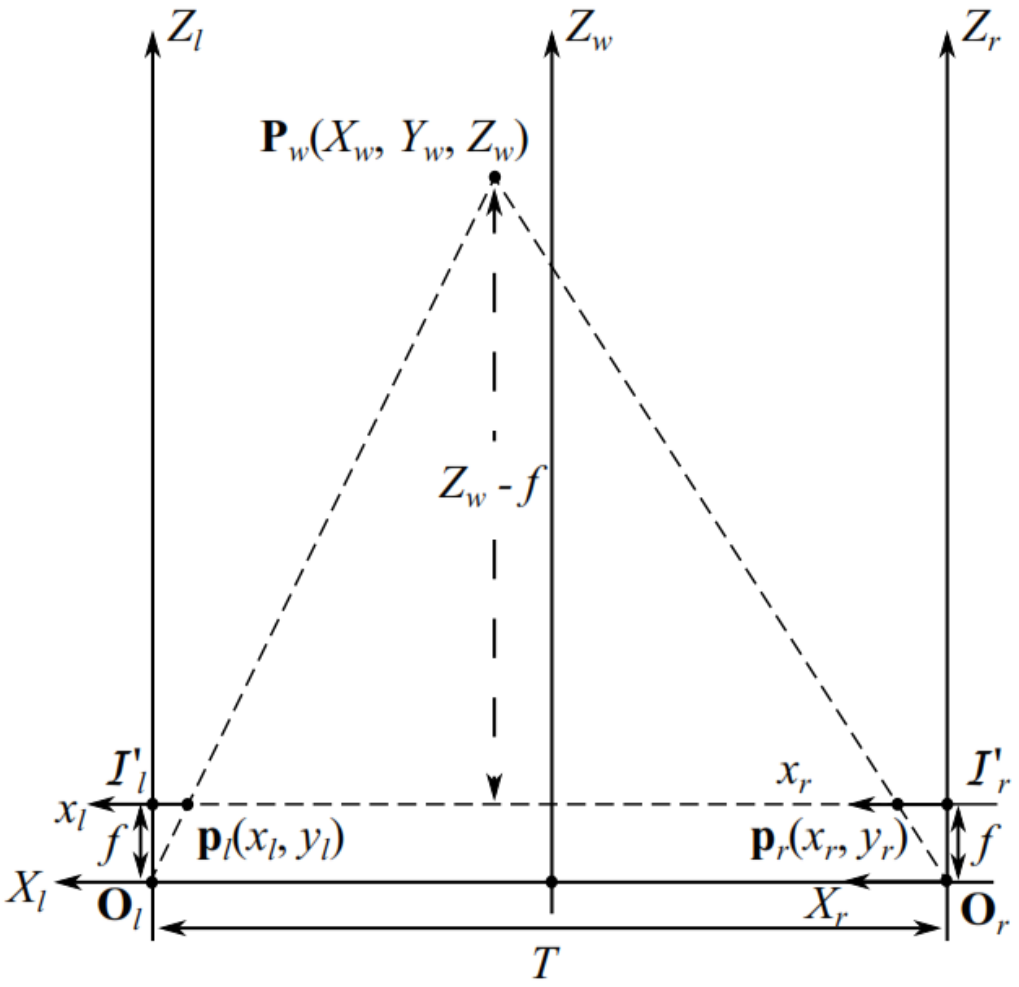
- Two cameras in known locations.
- Calibrated cameras.
- Stereo images can create a disparity (depth) map.

a) Left image; b) Right image; c) Dense disparity map.

Parallel Camera Setup

Parallel Stereo vision Geometry

T: baseline
f: focal length



Parallel Camera Setup

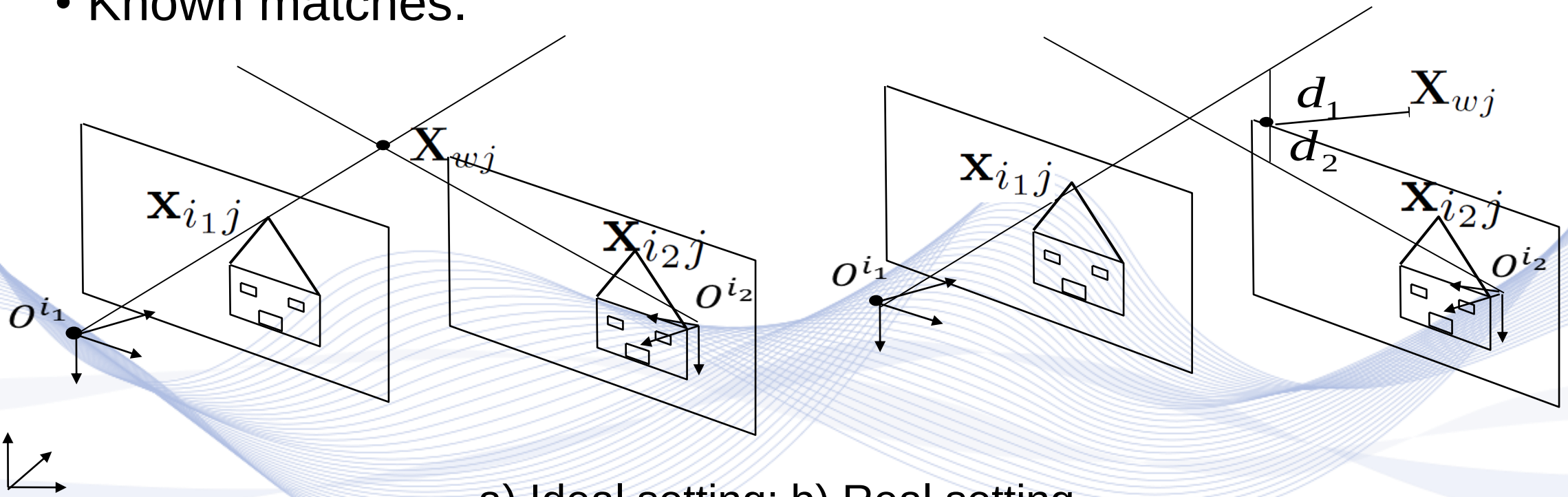
Triangle similarities can be used to recover 3D world coordinates $\mathbf{P}_w$ from left/right image plane coordinates, assuming all camera parameters and point disparity values $d_c = x_r - x_l$ are known:

$$Z_w = -\frac{fT_c}{d_c}, \qquad X_w = -\frac{T_c(x_l + x_r)}{2d_c}, \qquad Y_w = -\frac{T_c y_l}{d_c} = -\frac{T_c y_r}{d_c}.$$

- Camera calibration parameters must be known to refer to a real world coordinate system.

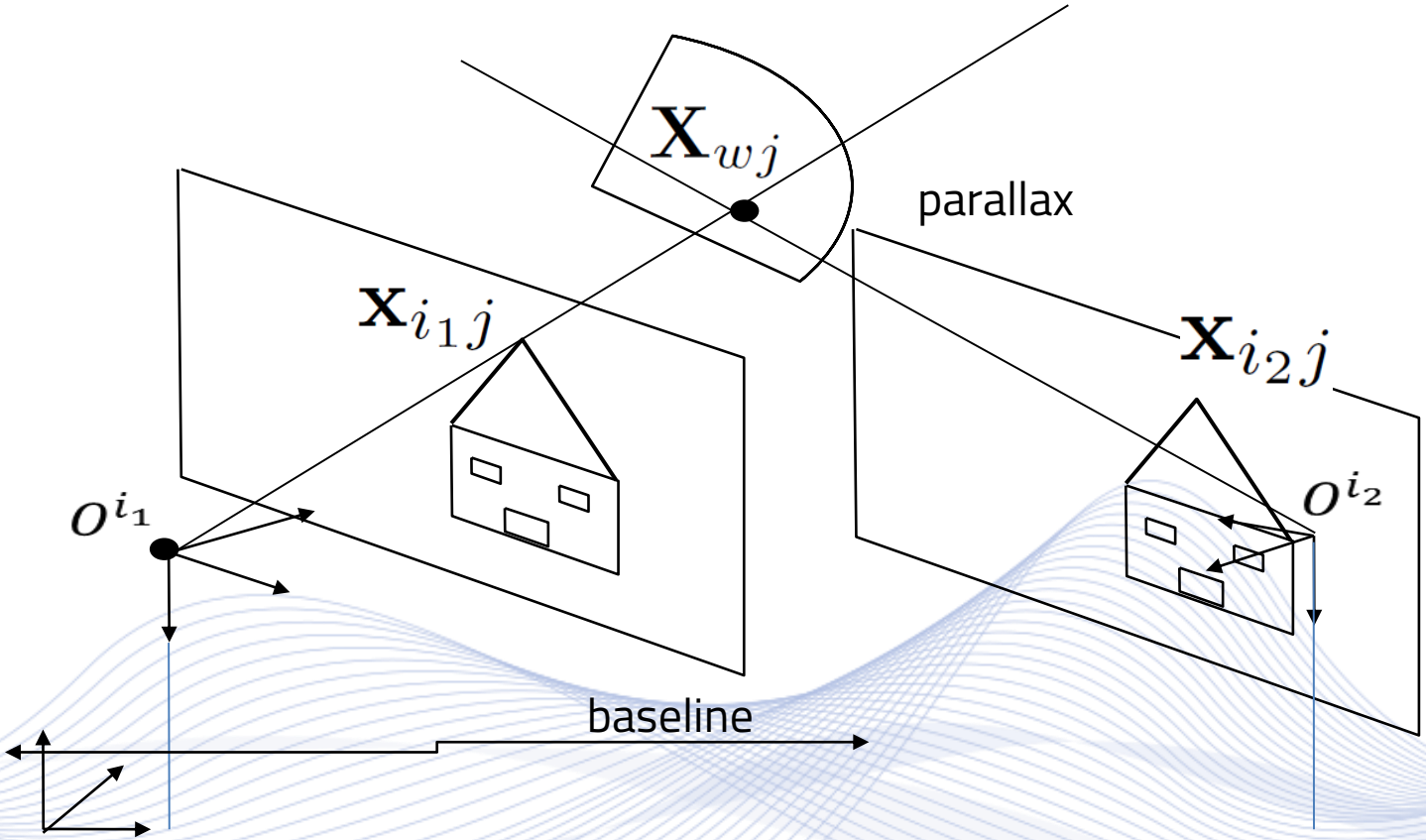
3D perception

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



a) Ideal setting; b) Real setting.

3D perception



Geometrical accuracy depency on parallax angle.

Structure from Motion

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

Structure from Motion

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

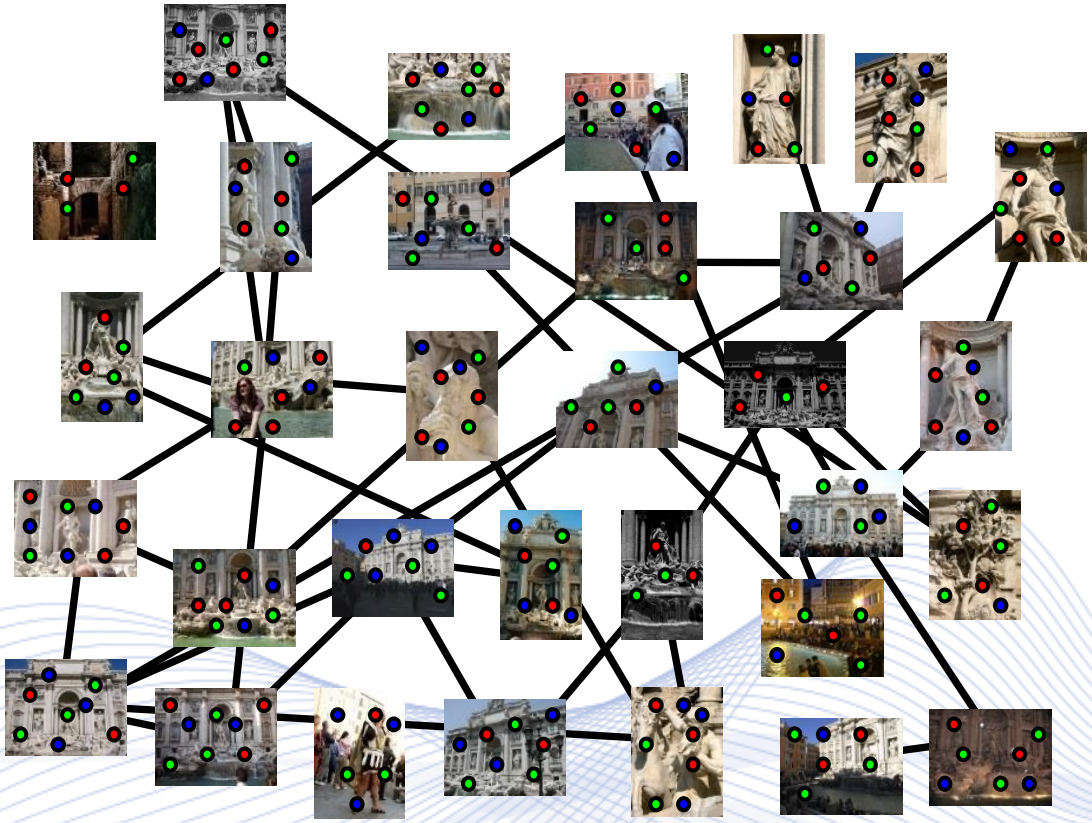


Photo tourism: exploring photo collections in 3D (https://www.youtube.com/watch?v=6eQ-CB8TY2Q)

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

- 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

Structure from Motion (SfM) perfoms two tasks simultaneously:

- 3D scene geometry reconstruction from a set of camera images and
- Camera calibration.

Images can be acquired by:

- multiple ***synchronized*** cameras or
- one moving camera, or unsyncronized multiple cameras, ***if the scene and illumination are static***.

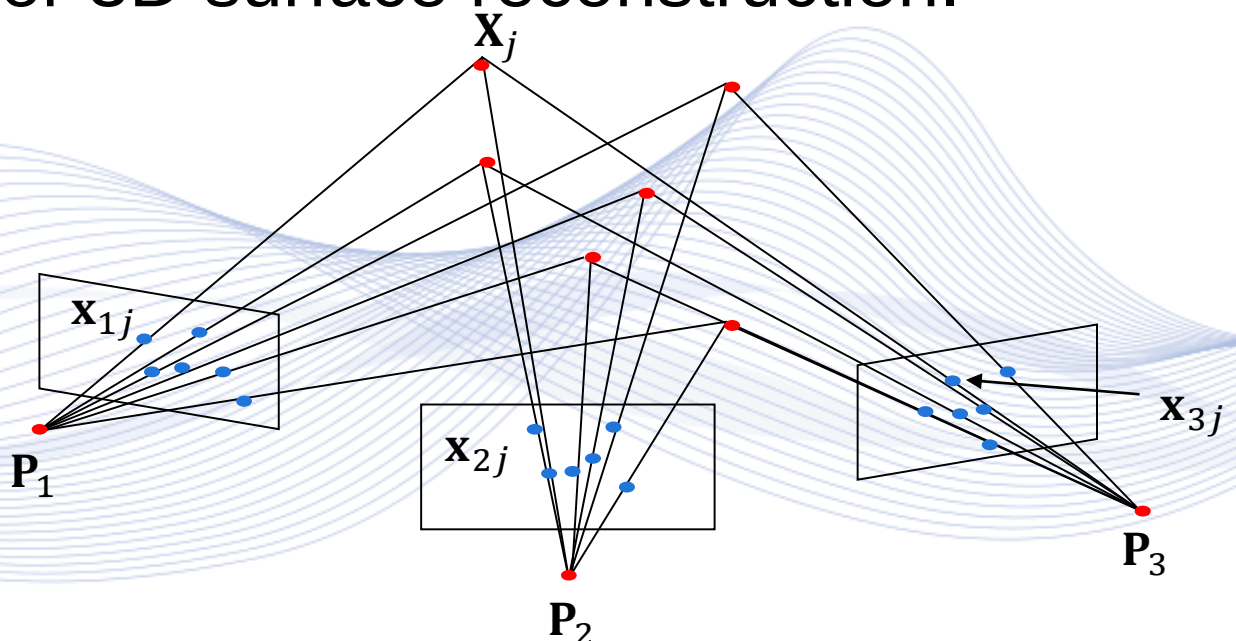
Feature extraction and matching

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

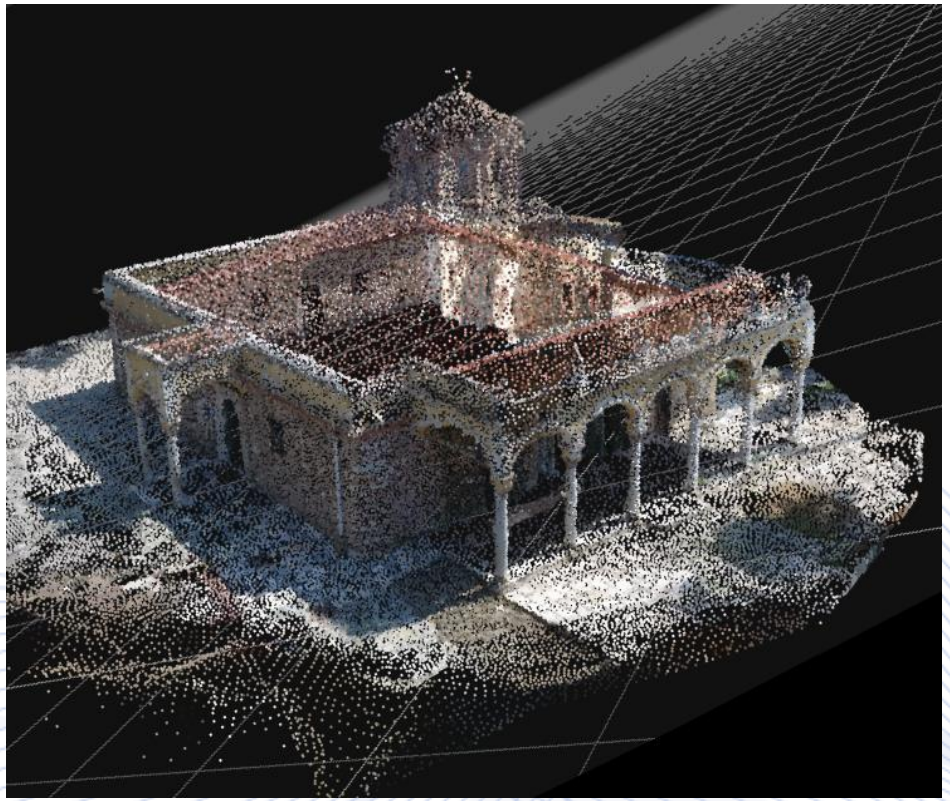
Vlatadon monastery [HEL2002].

Triangulation and Bundle Adjustment

- Bundle adjustment and triangulation are the final steps to estimate camera parameters and create an accurate point cloud.
- Further techniques are used to make the point cloud denser and to be used for 3D surface reconstruction.



Triangulation and Bundle Adjustment



3D model of Vlatadon monastery.

Camera Parameters and Projection Matrix

- Definition of the 3×4 matrix of extrinsic parameters $\mathbf{P}_E$:

$$\mathbf{P}_E = \begin{bmatrix} r_{11} & r_{12} & r_{13} & -\mathbf{R}_1^T\mathbf{T} \\ r_{21} & r_{22} & r_{23} & -\mathbf{R}_2^T\mathbf{T} \\ r_{31} & r_{32} & r_{33} & -\mathbf{R}_3^T\mathbf{T} \end{bmatrix}.$$

- Definition of the 3×3 matrix of intrinsic parameters $\mathbf{P}_I$:

$$\mathbf{P}_I = \begin{bmatrix} -\frac{f}{s_x} & 0 & o_x \\ 0 & -\frac{f}{s_y} & o_y \\ 0 & 0 & 1 \end{bmatrix}.$$

Camera Parameters and Projection Matrix

The transformation of a point $\mathbf{P} \in \mathbb{P}^3$ to $\mathbf{p} \in \mathbb{P}^2$ is given by:

$$\begin{bmatrix} Zx_d \\ Zy_d \\ Z \end{bmatrix} = \mathbf{P}_I \mathbf{P}_E \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}.$$

$$\mathbf{p} = \mathbf{P}_I \mathbf{P}_E \mathbf{P} = \boldsymbol{\mathcal{P}} \mathbf{P}.$$

Structure from Motion

Bundle Adjustment:

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

$$\underset{\boldsymbol{\mathcal{P}}_i, \mathbf{P}_j}{\operatorname{argmin}} \sum_{i,j} v_{ij} \left\| \mathbf{p}_{ij} - \boldsymbol{\mathcal{P}}_i \mathbf{P}_j \right\|^2.$$

- $\boldsymbol{\mathcal{P}}_i$: projection matrix of camera i.
- $\mathbf{P}_j$: world coordinate point $\mathbf{X}_j$ (homogeneous coordinates).
- $\mathbf{p}_{ij}$: projection $\mathbf{x}_{ij}$ on camera i plane (homogeneous coordinates).
- $v_{ij} = \{0,1\}$: it denotes if point j is visible on camera i.

Structure from Motion

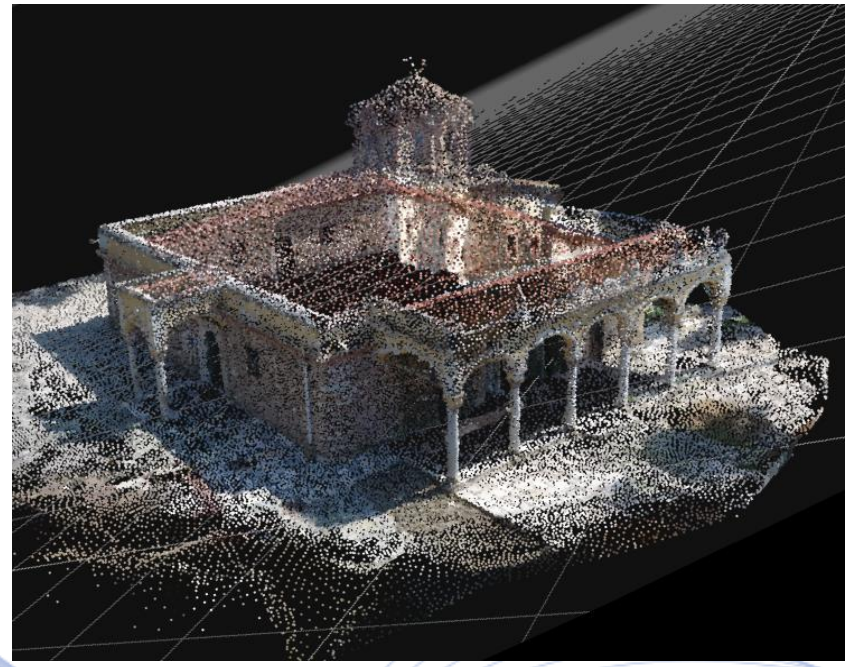
- Minimization of the reprojection error.
- Nonlinear least-squares minimization algorithms.
- ***Levenberg–Marquardt*** is the most successful.
- Iterative error function linearization in a local neighborhood of the current estimate, leads to solving the linear ***normal equations***.
- They have sparse block structure, leading to very fast algorithms.

Structure from Motion

- A 3D point must be visible from at least 4 cameras.
- The more views we have, the better SfM results.

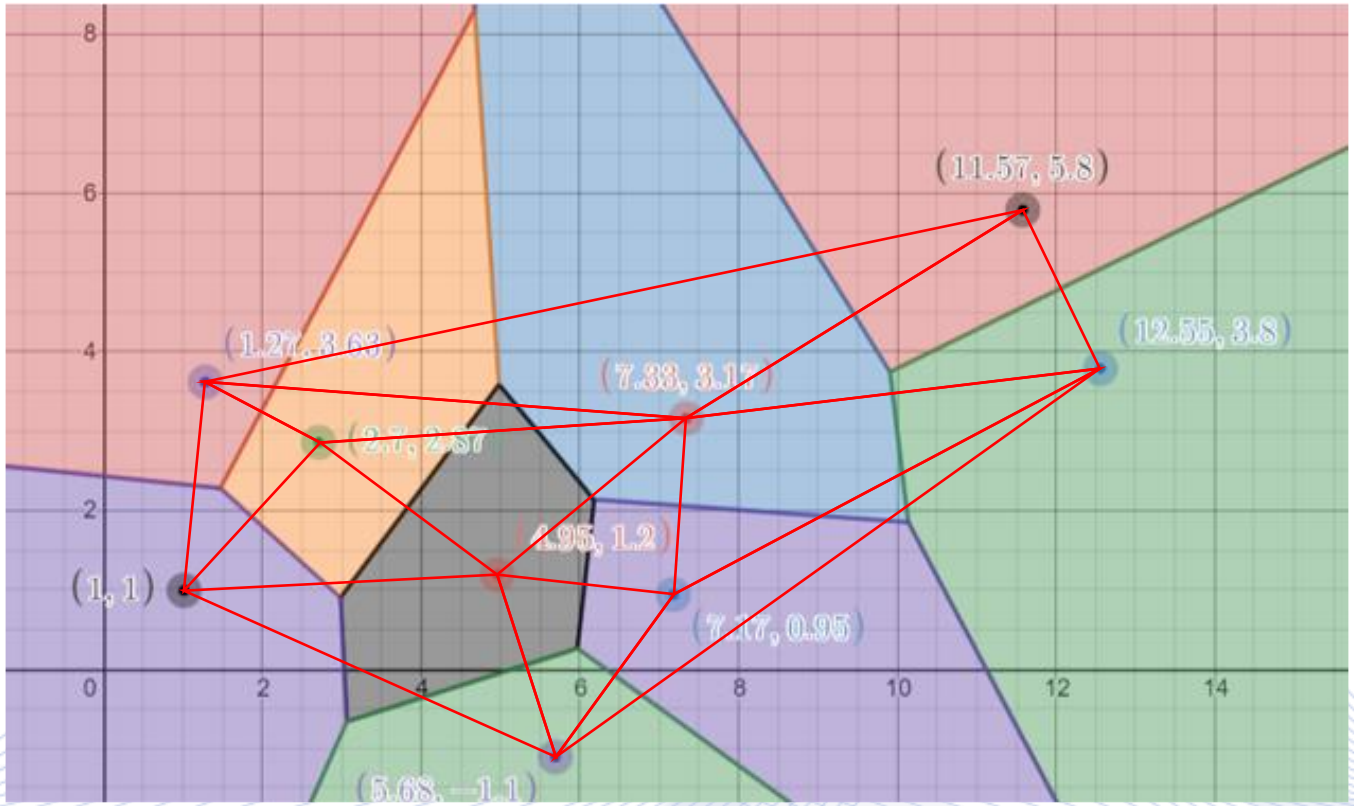
- Sensitivity to noise and outliers:
  - Moving objects.

Structure from Motion



Polygonal surface mesh.

Structure from Motion

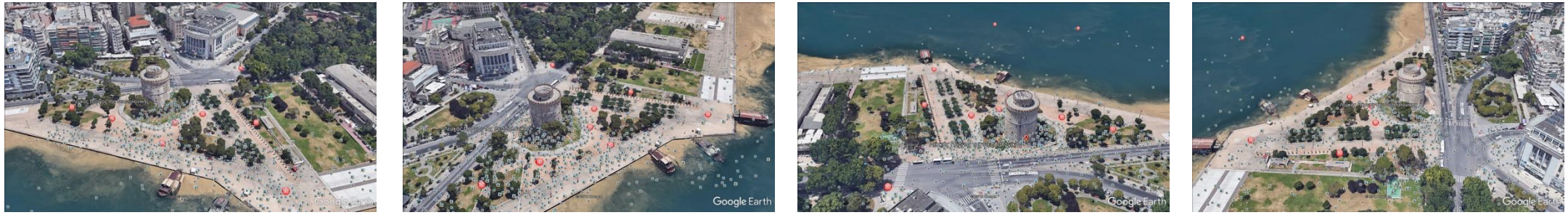


Voronoi tessellation and Delaunay triangulation.

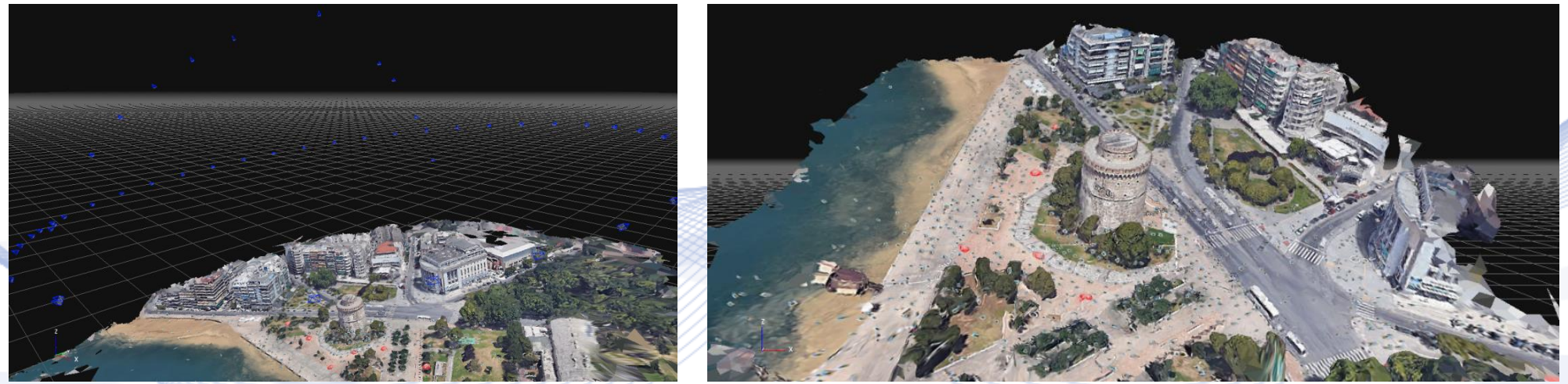
Structure from Motion

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

Structure from Motion



Images obtained from Google Earth.



3D reconstructed model using 50 images from Google Earth.

SfM in 3D Face Reconstruction

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

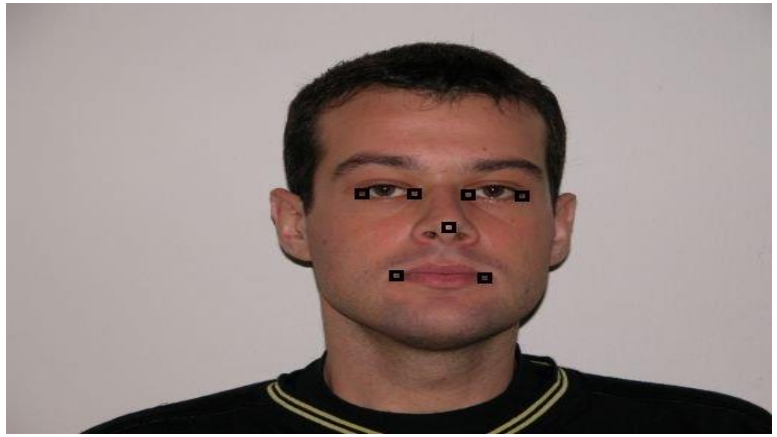
3D face reconstruction, facial pose estimation, face recognition, face verification.

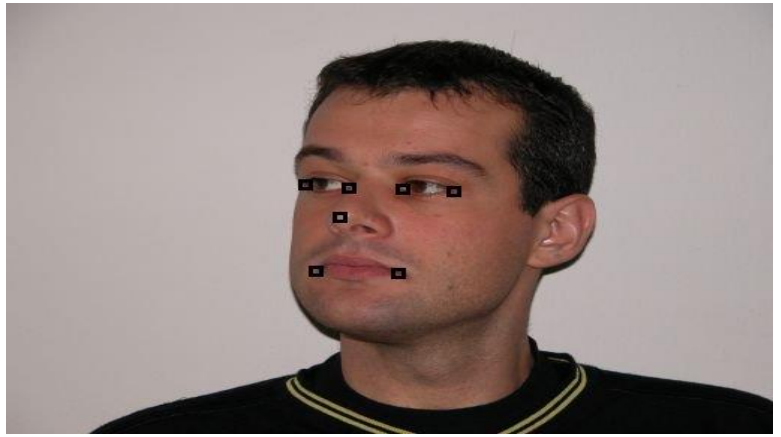
SfM in 3D Face Reconstruction

Method overview

- Manual selection of characteristic feature points on the input facial images.
- Use of an uncalibrated 3D reconstruction algorithm.
- Incorporation of the CANDIDE generic face model.
- Deformation of the generic face model based on the 3-D reconstructed feature points.
- Re-projection of the face model grid onto the images and manual refinement.

SfM in 3D Face Reconstruction





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

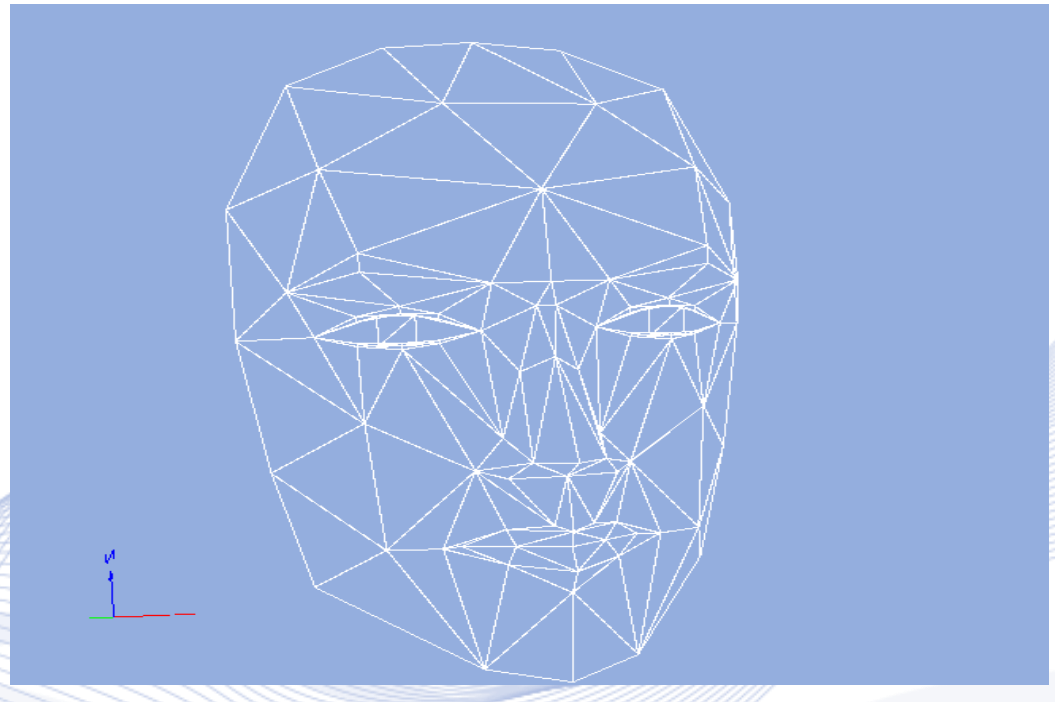
SfM in 3D Face Reconstruction

Having established feature point correspondences, 3D geometry reconstruction is performed:

- Calculation of the epipolar geometry based on the first two images of the input set.
- Derivation of the first two projection matrices.
- Calculation of the first depth estimate.
- Incorporation of the rest of the images of the input set.
- Bundle Adjustment (global refinement).
- Camera self calibration.

SfM in 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

- Deformation of the generic face model by a ***Finite Element Method*** (***FEM***).

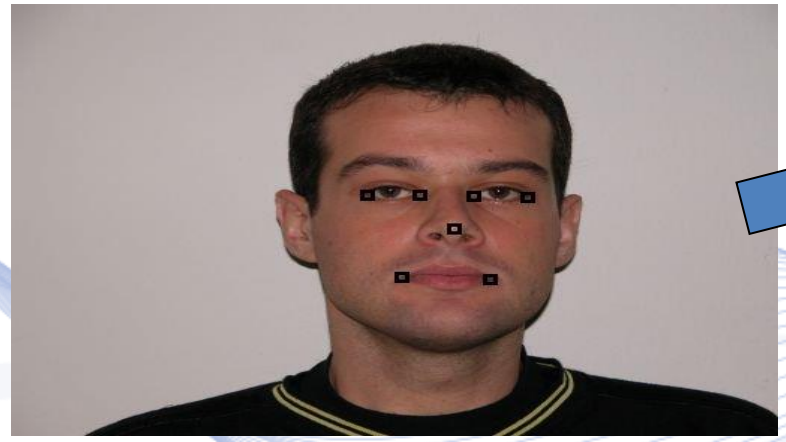
- Input to FEM: the list of the reconstructed 3D face points and the corresponding CANDIDE nodes.

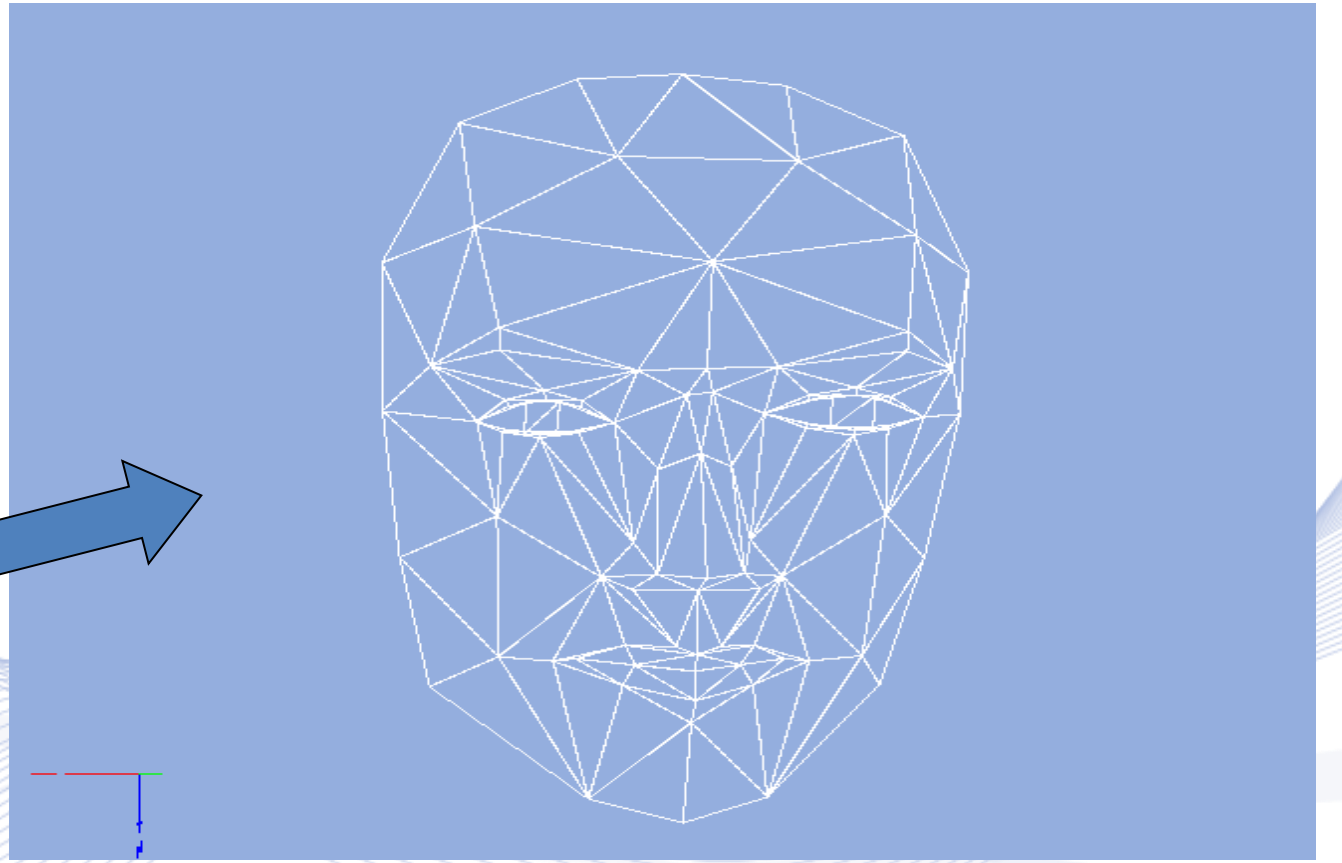
- the CANDIDE model is scaled, rotated and translated in order to be aligned roughly with the reconstructed 3-D feature points.

- Finally, the CANDIDE model is deformed locally, in order to fit the specific face characteristics.

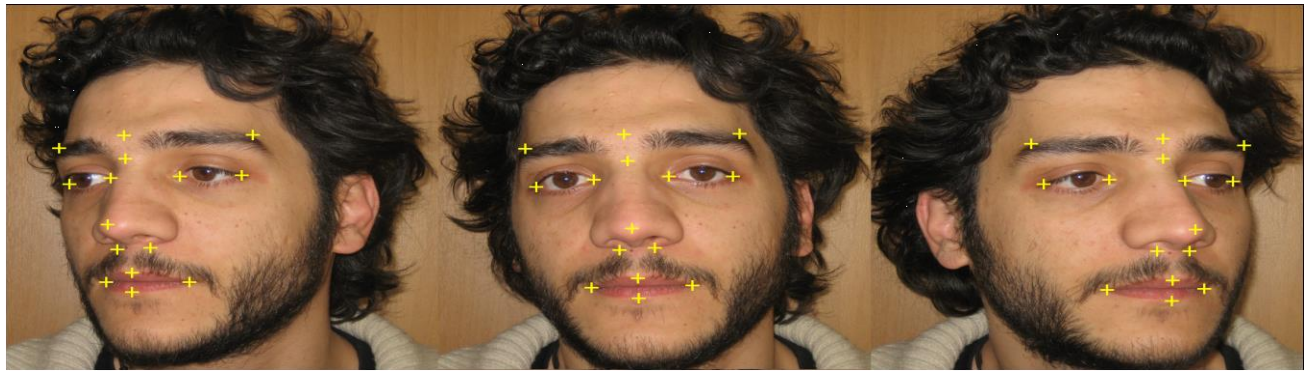
SfM in 3D Face Reconstruction

Example of Candide fitting to facial images.





SfM in 3D Face Reconstruction



Selected features.

CANDIDE grid reprojection.

Reconstructed 3D face model.

SfM in 3D Face Reconstruction

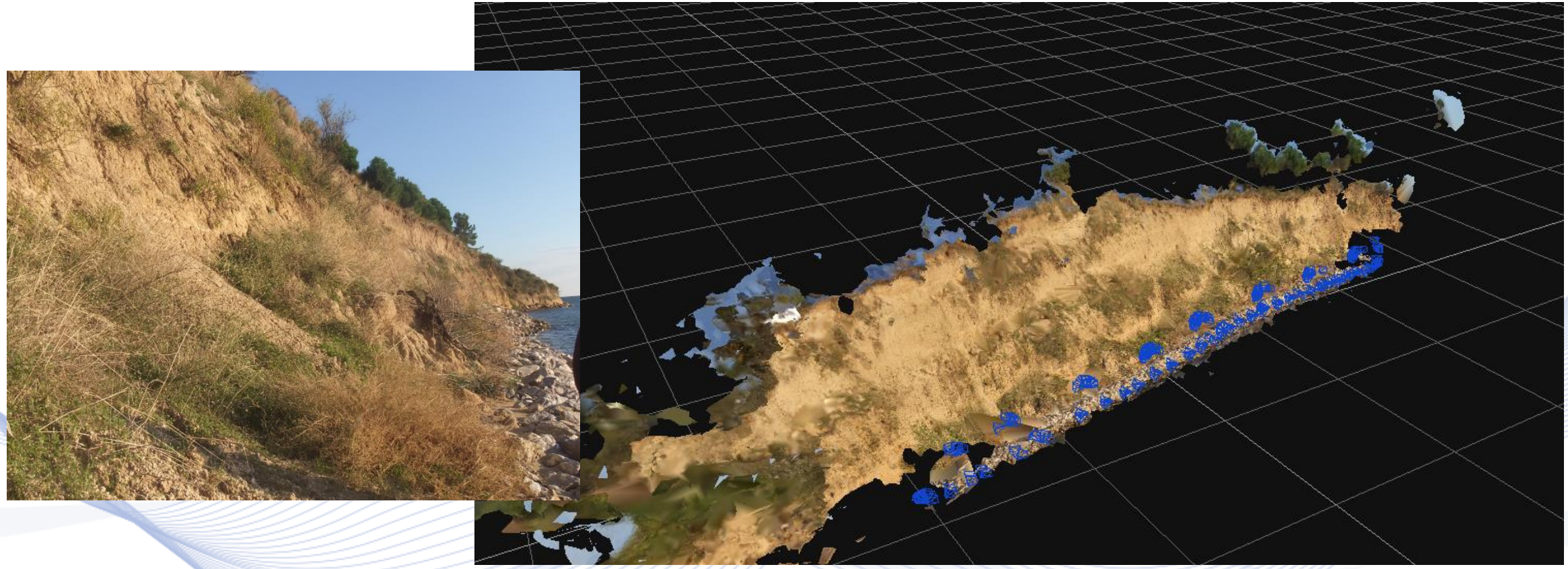
Performance evaluation: Reprojection error.

Manually Selected Coordinates	Calculated Coordinates	Reprojection Error (pixels)
(1131,1151)	(1131,1151)	(0,0)
(1420,1164)	(1420,1164)	(0,0)
(1050,776)	(1051,775)	(-1,1)
(1221,786)	(1218,788)	(3,-2)
(1392,794)	(1395,795)	(-3,-1)
(1567,793)	(1566,792)	(1,1)
(1244,957)	(1244,957)	(0,0)

SfM in 3D landscape reconstruction

Cliff images.

SfM in 3D landscape reconstruction



3D cliff surface reconstruction.

SfM in 3D monument reconstruction

- Two of the fifteen Paleochristian and Byzantine monuments of Thessaloniki that were included in the UNESCO World Heritage List:

  - Vlatadon Monastery;
  - Church of Saint Nicholas Orphanos.

SfM in 3D monument reconstruction

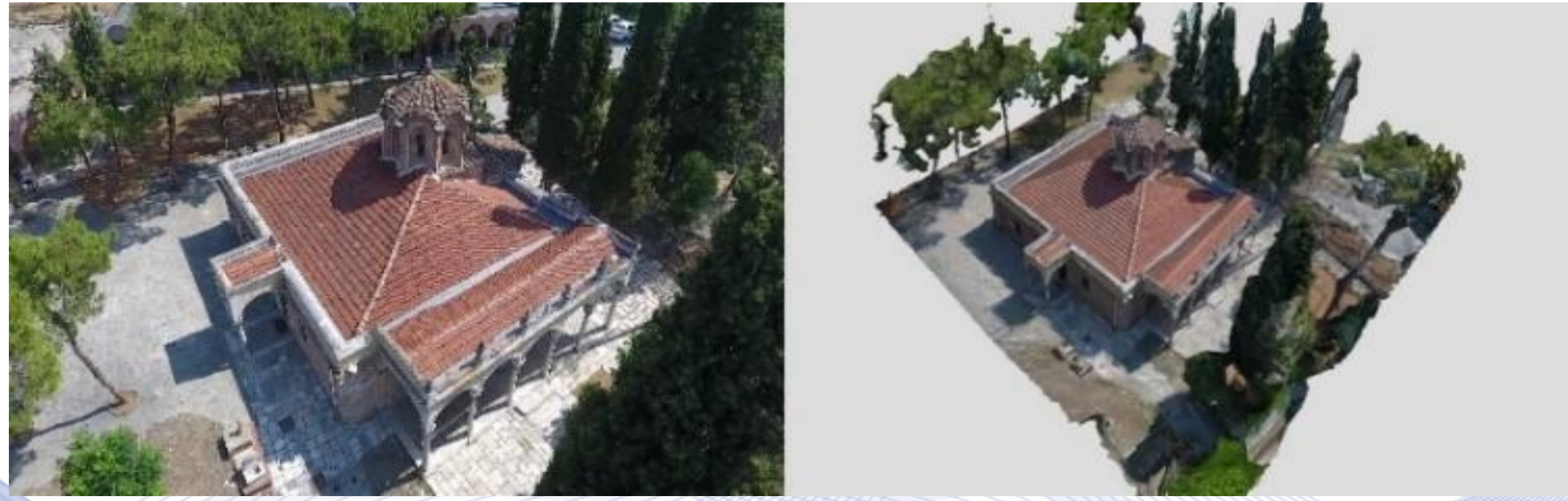
- Imaging by mostly orbiting a drone around them at different heights, respecting always the corresponding flight regulations:
  - avoiding collisions with nearby objects (trees, buildings, etc.) had a negative impact in capturing the close-by details of each monument.
  - Additional image collection from ground cameras.

- ***3DF Zephyr***: a commercial software for SoA 3D scene reconstruction.

SfM in 3D monument reconstruction

3 views of Vlatadon monastery, Thessaloniki, Greece [HEL2002].

SfM in 3D monument reconstruction



a) Real image; b) Synthetic 3D model view of Vlatadon Monastery.

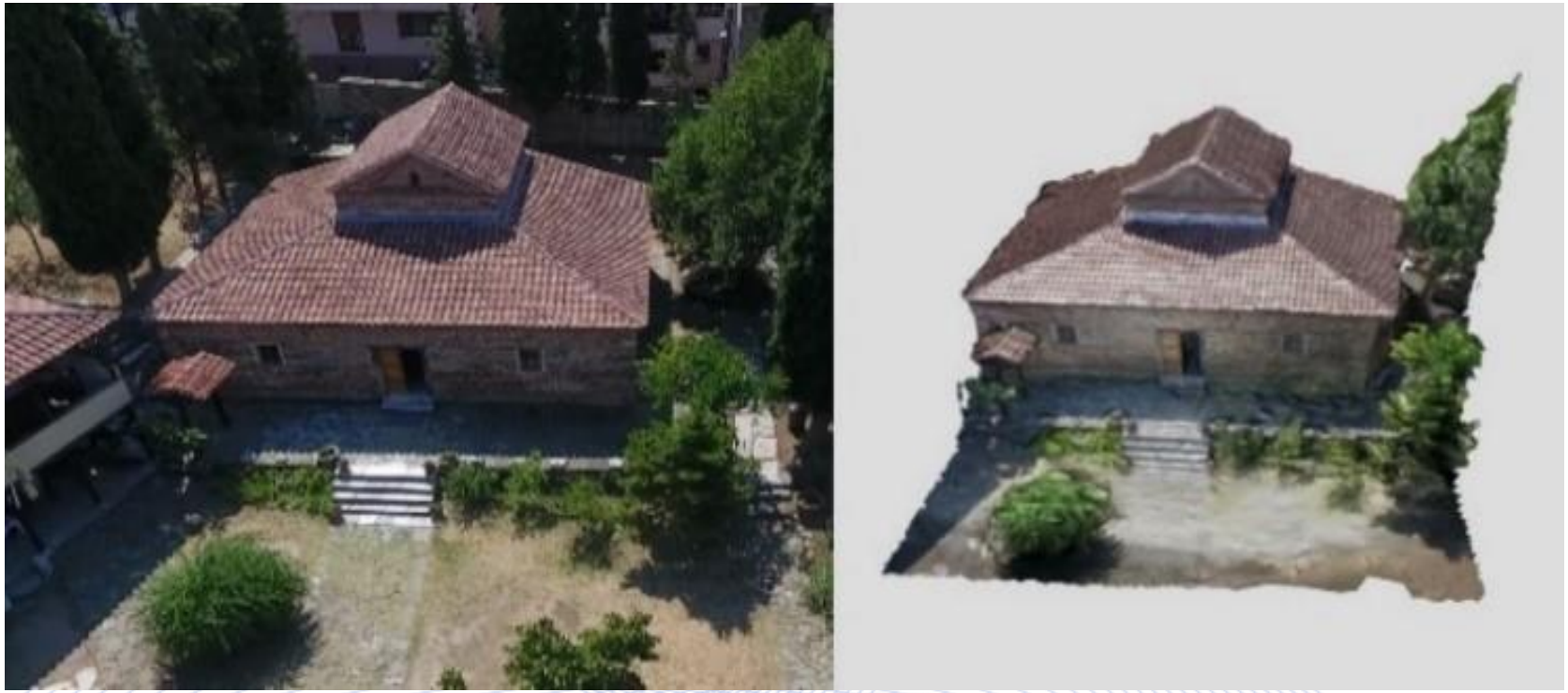
SfM in 3D monument reconstruction

3D model of Vlatadon monastery.

SfM in 3D monument reconstruction

St Nicolas Orphanos church, Thessaloniki, Greece [HEL2002].

SfM in 3D monument reconstruction

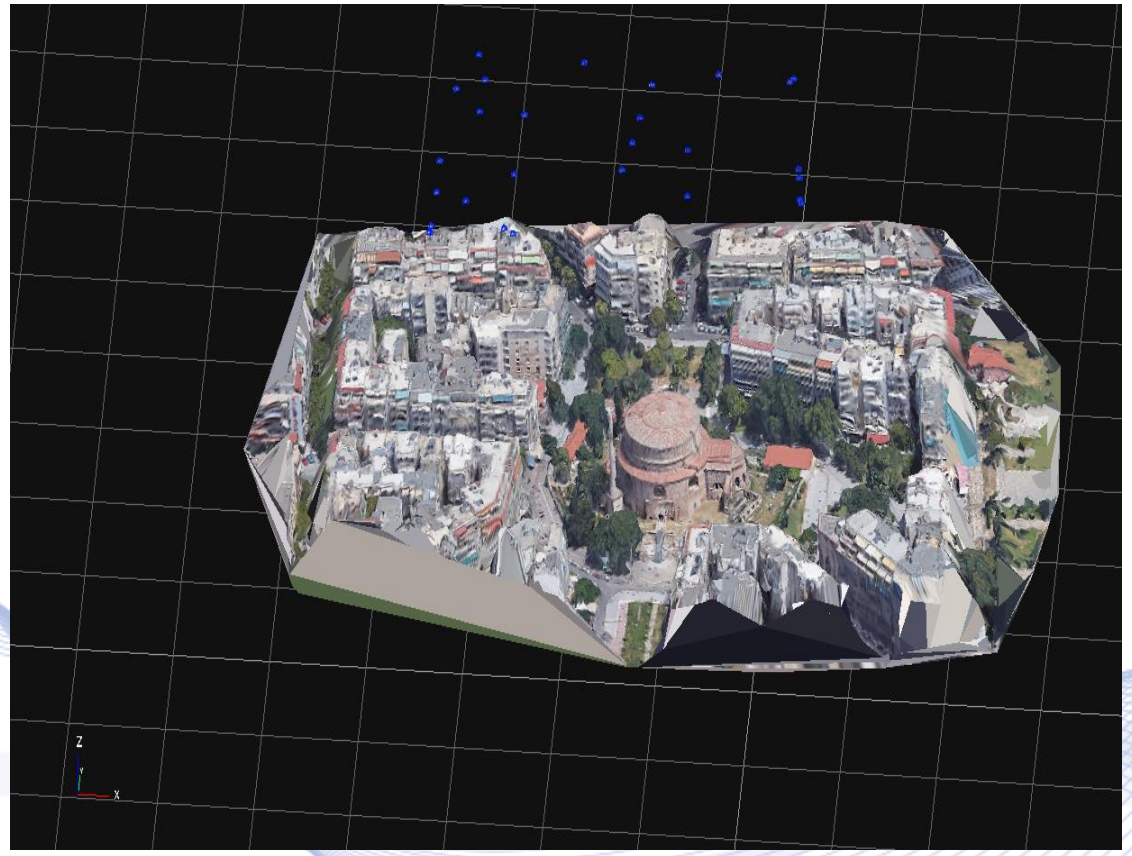


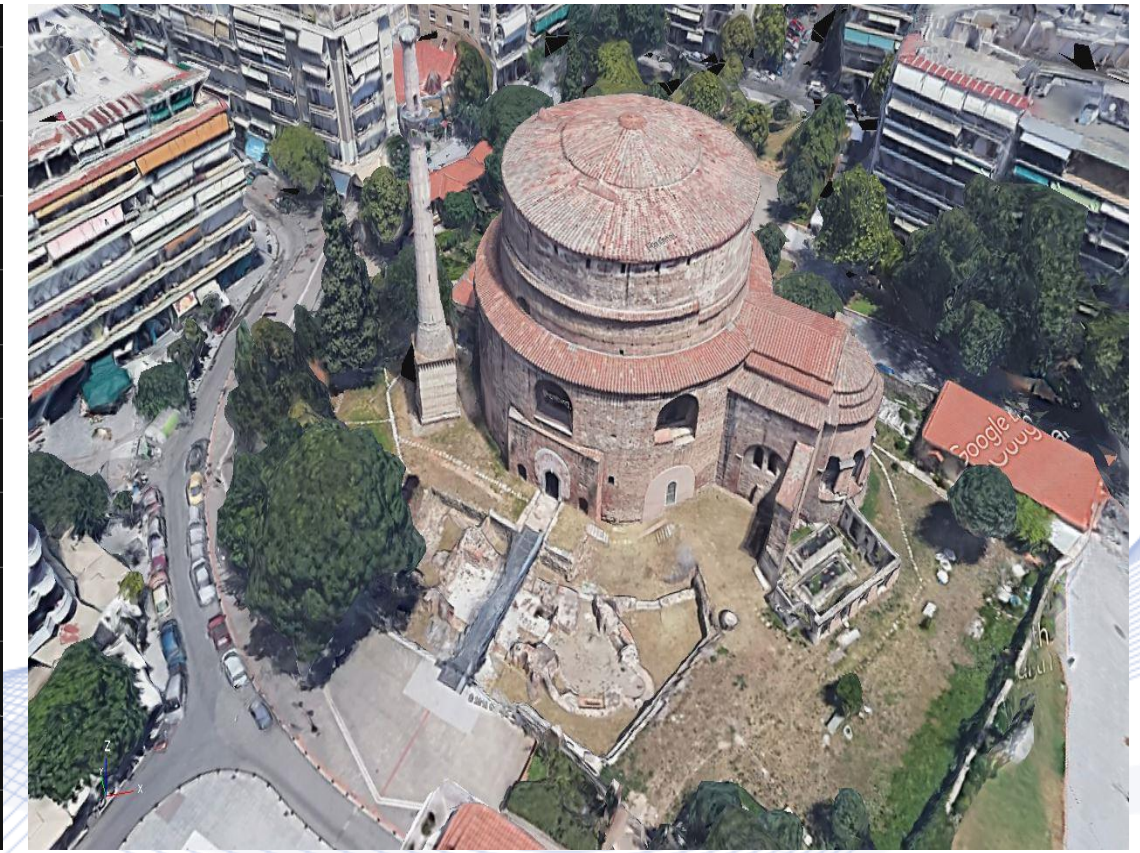
a) Real image; b) Synthetic 3D model view of St Nicolas Orphanos church.

SfM in 3D monument reconstruction

- 3D reconstruction using images from Google Maps/Earth.
- Rotunda: a Paleochristian monument of Thessaloniki, included in the UNESCO World Heritage List.
- 3D reconstruction of Rotunda with relative low detail level, using minimum resources, without image collection neither from the ground nor from drones.
- Images of the monument publicly available on the Internet (e.g. Google Maps/Earth) were imported.
- 3DF Zephyr: a commercial software for SoA 3D scene reconstruction.

SfM in 3D monument reconstruction





Rotonda monument, Thessaloniki, Greece.

Structure from Motion

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

UAV Image Capturing

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.

UAV Image Capturing

Scanning Strategy

- Information required:
  - Location of the environmental area.
  - Camera parameters (sensor size and focal length).

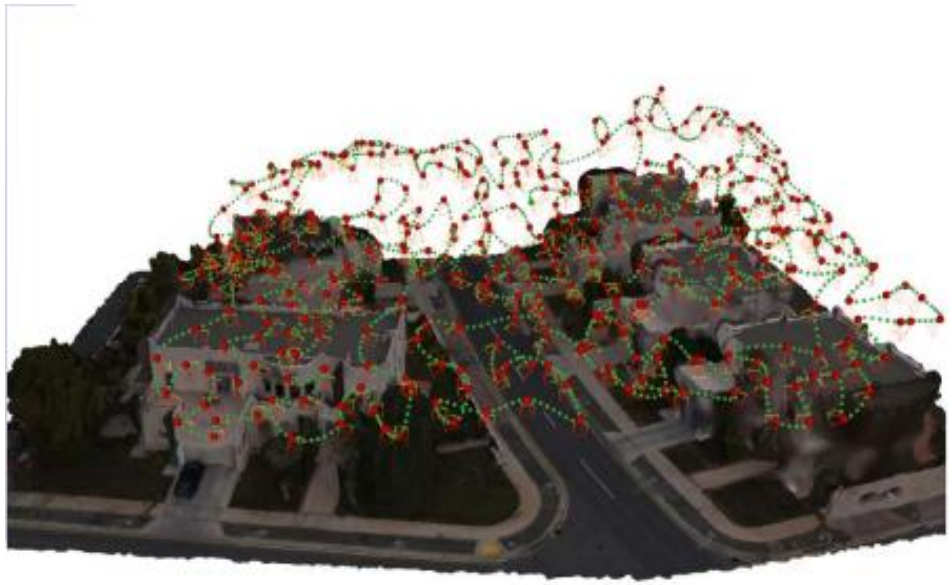
- Optimal shot parameters:
  - Flight trajectory
  - Flying heights
  - Viewing angles
  - Image overlap ratios (the number of images).

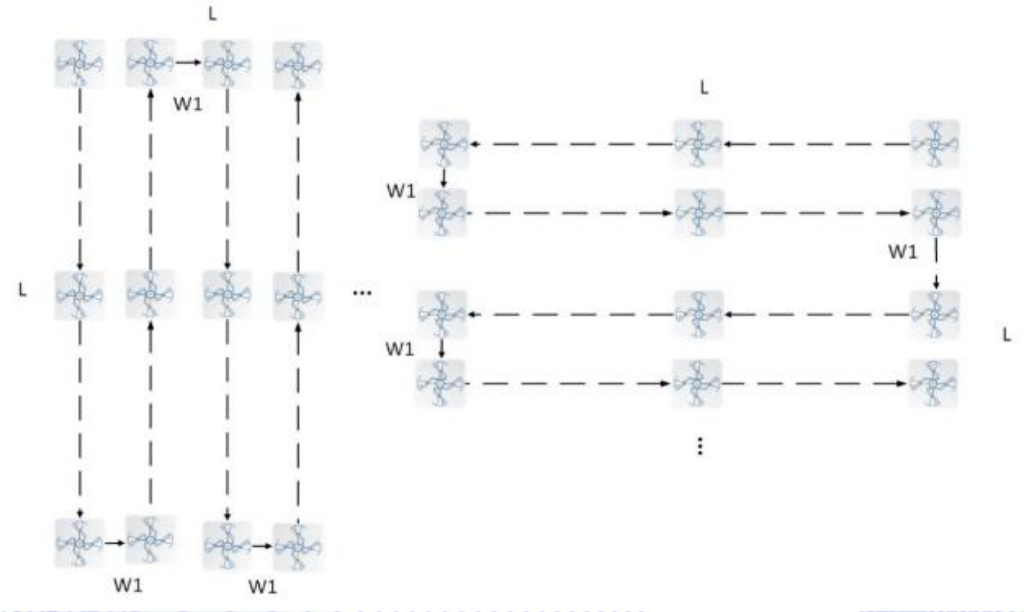
Optimal UAV Flight Trajectory

- The optimal flight trajectory for a specific background environment ***depents highly on the given landscape and object complexity*** [SMI2018].

- 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.

Optimal UAV Flight Trajectory





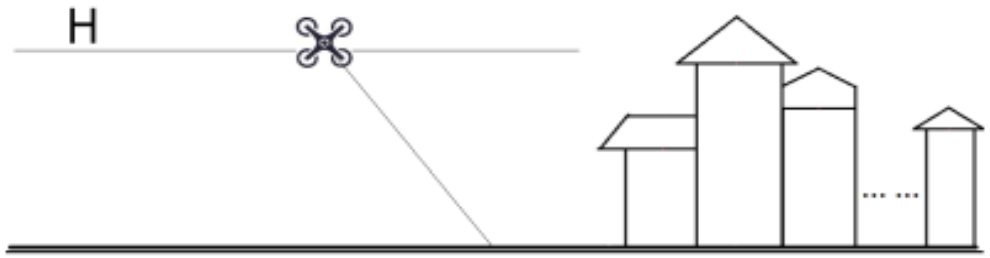
UAV Flight Height

- A ***single layer of scanning height*** was recommended in [SEI2019] for a plain landscape scenario, using a fixed height value.
- Based on the recommendation in [HAW2016], a ***three layer scanning approach*** can be employed to generalize the height configuration for both landscape and urban environments:

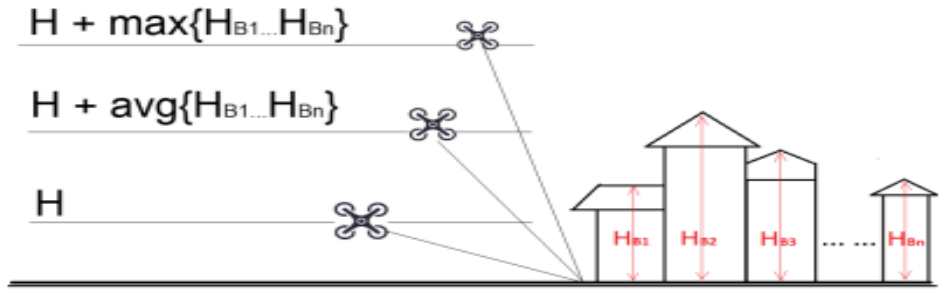
$$\text{Drone Flight Height} = \begin{cases} H, & 1^{\text{st}} \text{ level;} \\ H + \text{average building height}, & 2^{\text{nd}} \text{ level;} \\ H + \text{maximum building height}, & 3^{\text{rd}} \text{ level.} \end{cases}$$

UAV Flight Height

20 m is used for H that is within the recommended range in [SEI2019] for a default camera sensor size of 23.66×13.3 mm and a focal length of 35 mm.



(a) Single layer of height in [Seifert *et al.*, 2019].



(b) Three layers of scanning height employed.

UAV Camera Viewing Angles

- ***Viewing angle*** is also called ***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 degrees.

- Other parameters such as flight heights and the number of images are fixed.

UAV Camera Viewing Angles

- 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

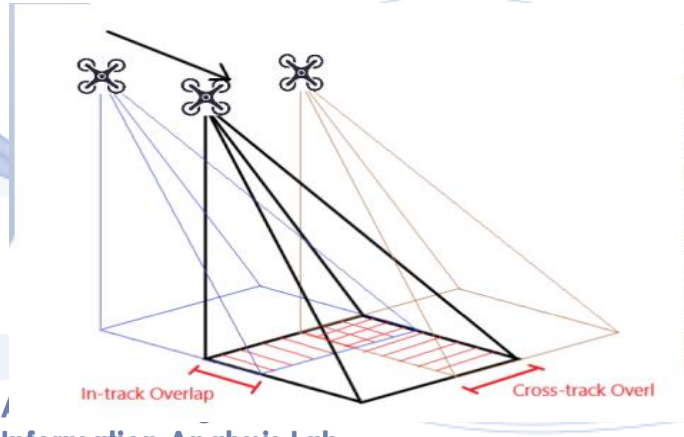
Image Overlap Ratio

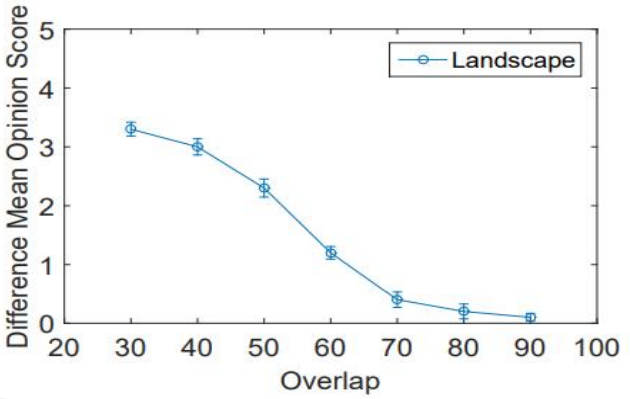
- ***In-track overlap***: between adjacent images captured in the same flight track.
- ***Cross-track overlap***: between the adjacent images captured at neighboring flight tracks.

- Seven in-track overlap values have been employed to capture images in the Country Side scenario (from UE4 market place) from ***30% to 90%*.**

Image Overlap Ratio

- The same flight height and trajectories, viewing angles and cross-track overlap were employed here.
- A ***subjective test*** was conducted to evaluate the influence of in-track overlap on reconstruction quality.
- Based on the results, we recommend to use 70% and 80% overlap ratios for scenarios of landscapes and detailed objects, respectively.





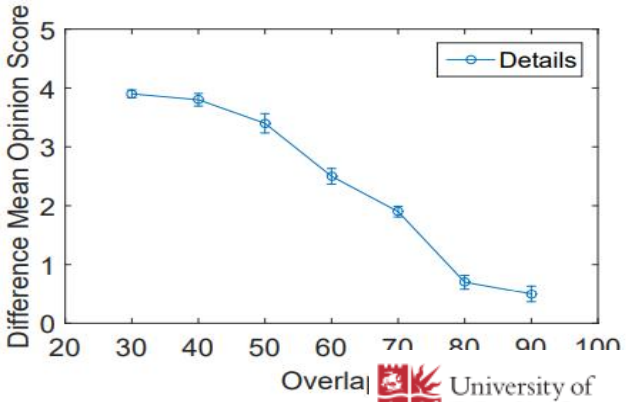
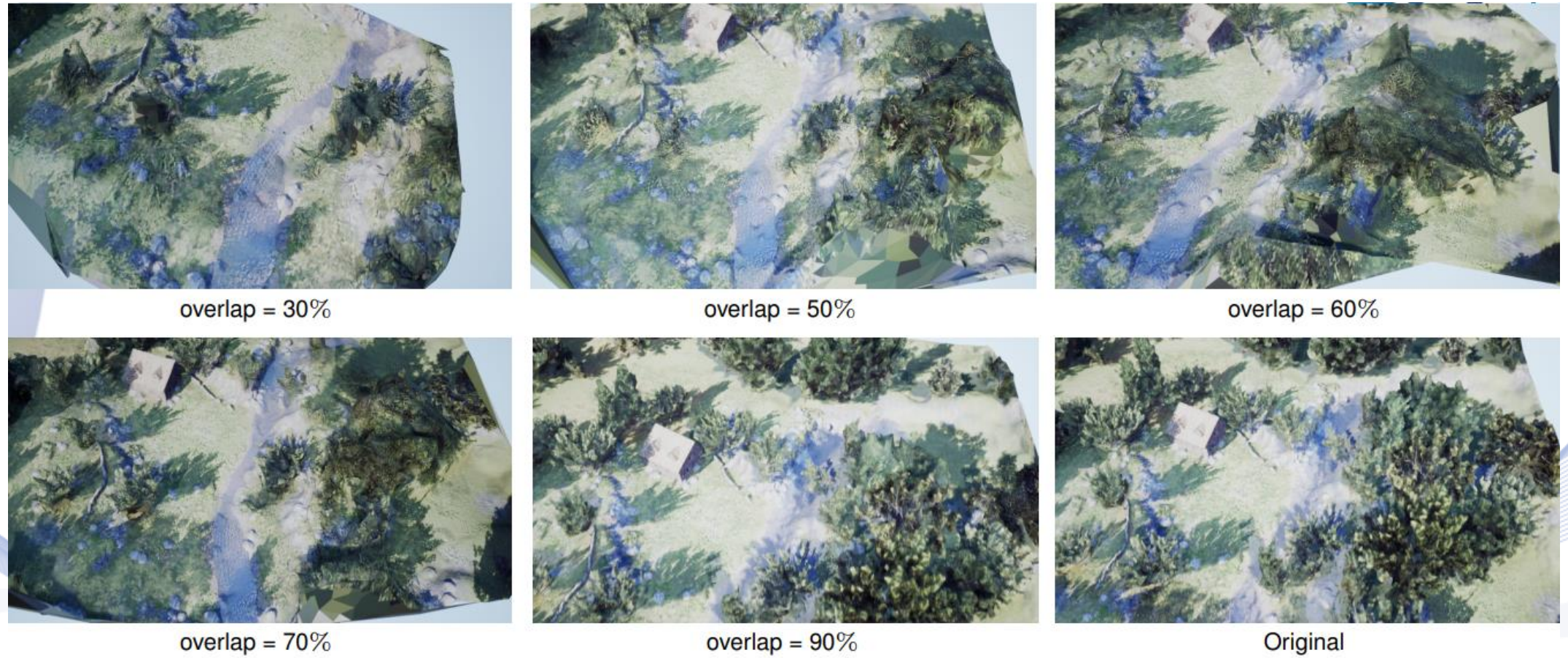


Image Overlap Ratio

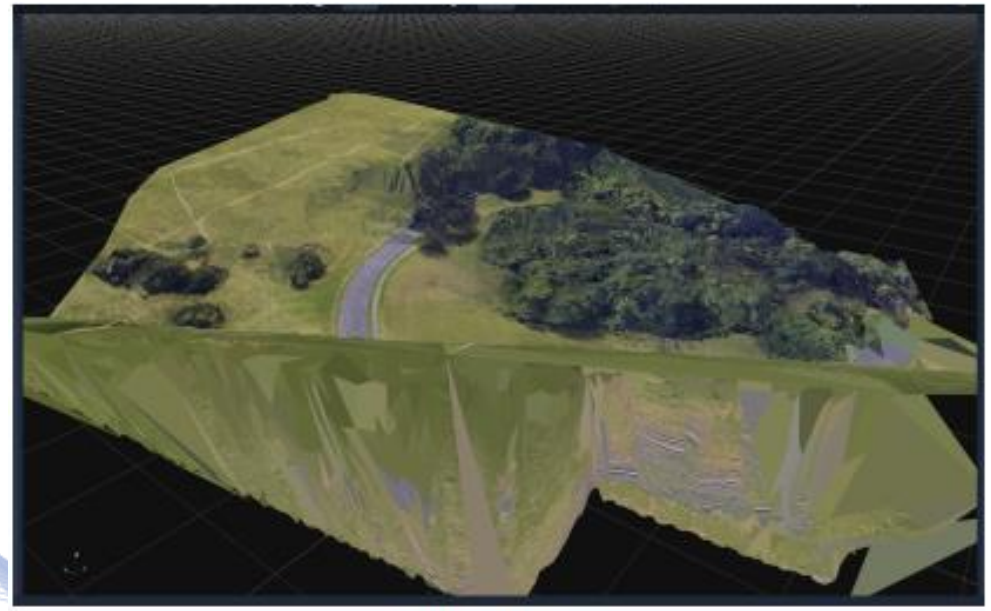


Photogrammetry Reconstruction

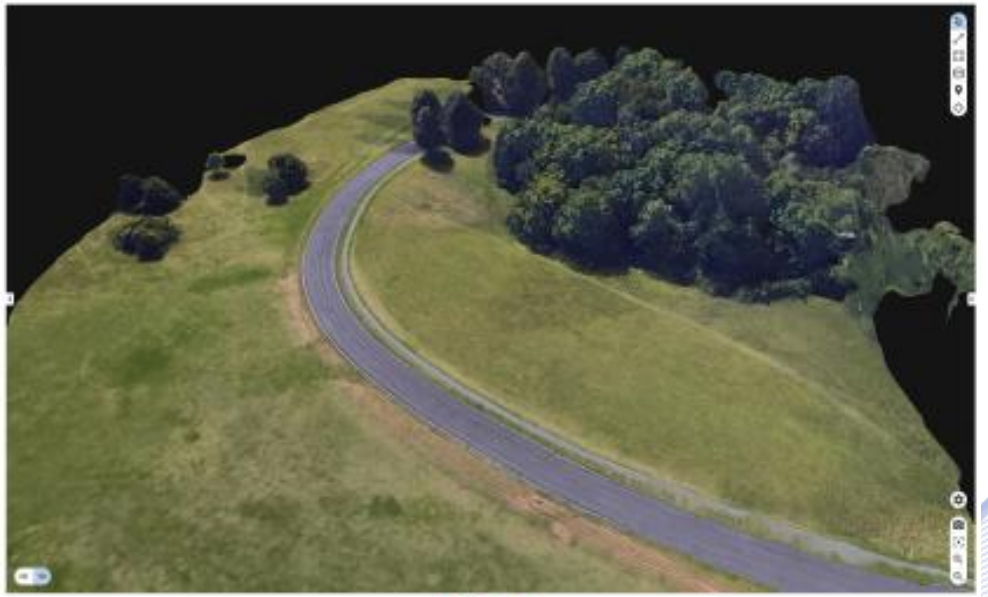
3D photogrammetry software packages

- 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).

Photogrammetry Reconstruction



3DF Zephyr



Pix4D

Pre-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.
- Such artefacts can be corrected through ***texture inpainting*** [TSC2005], [WON2006] .
- Most of existing in-painting algorithms are relatively complex computationally expensive.
- Simple manual outlier rejection can be applied on the input image dataset in order to achieve efficient reconstruction.

Post-processing for 3D Reconstruction

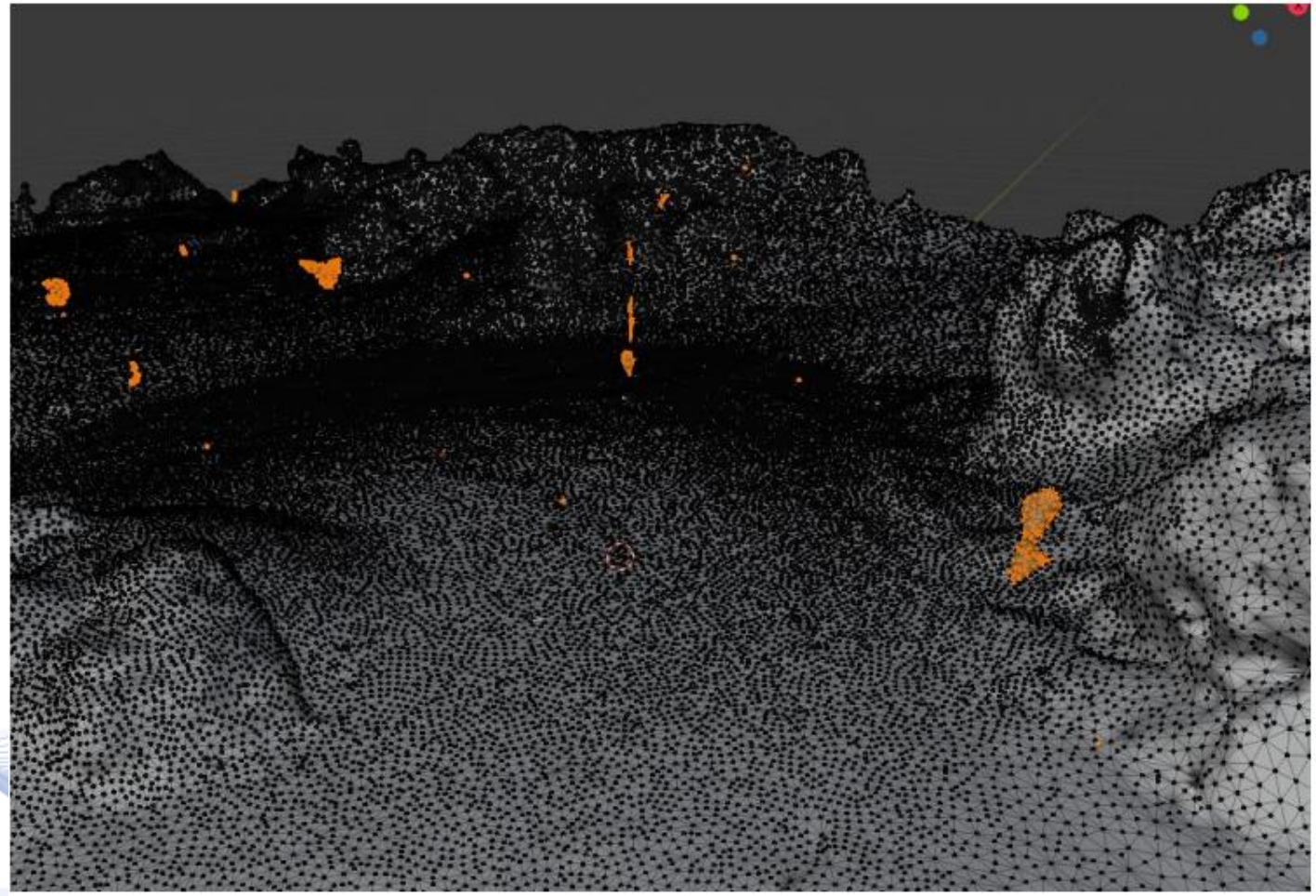
Post-processing:

- After pre-processing, photogrammetry software can provide reasonably good results.
- There are still a large amount of noticeable artefacts, which could affect viewing experience.
  - Bumps and holes.
- 3D model editing can be employed to further correct these distortions.

Post-processing for 3D Reconstruction

- 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

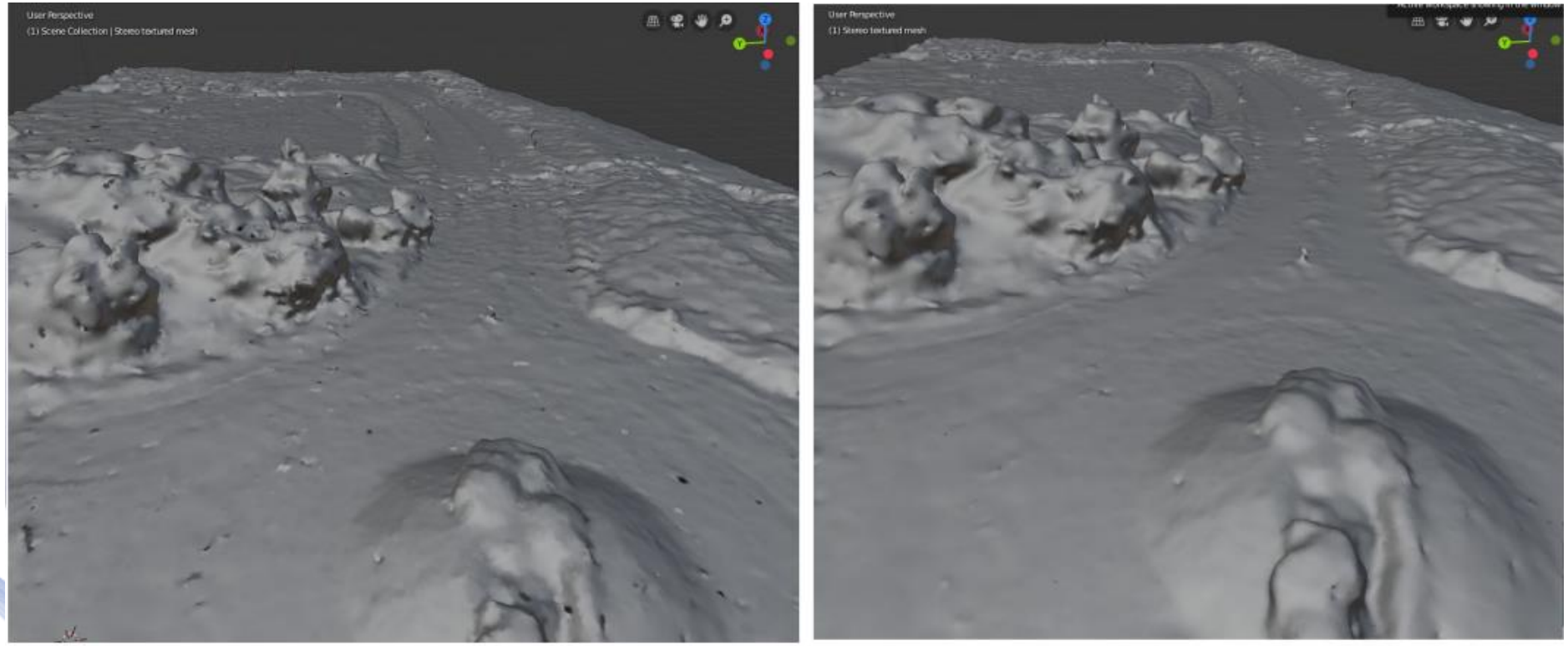
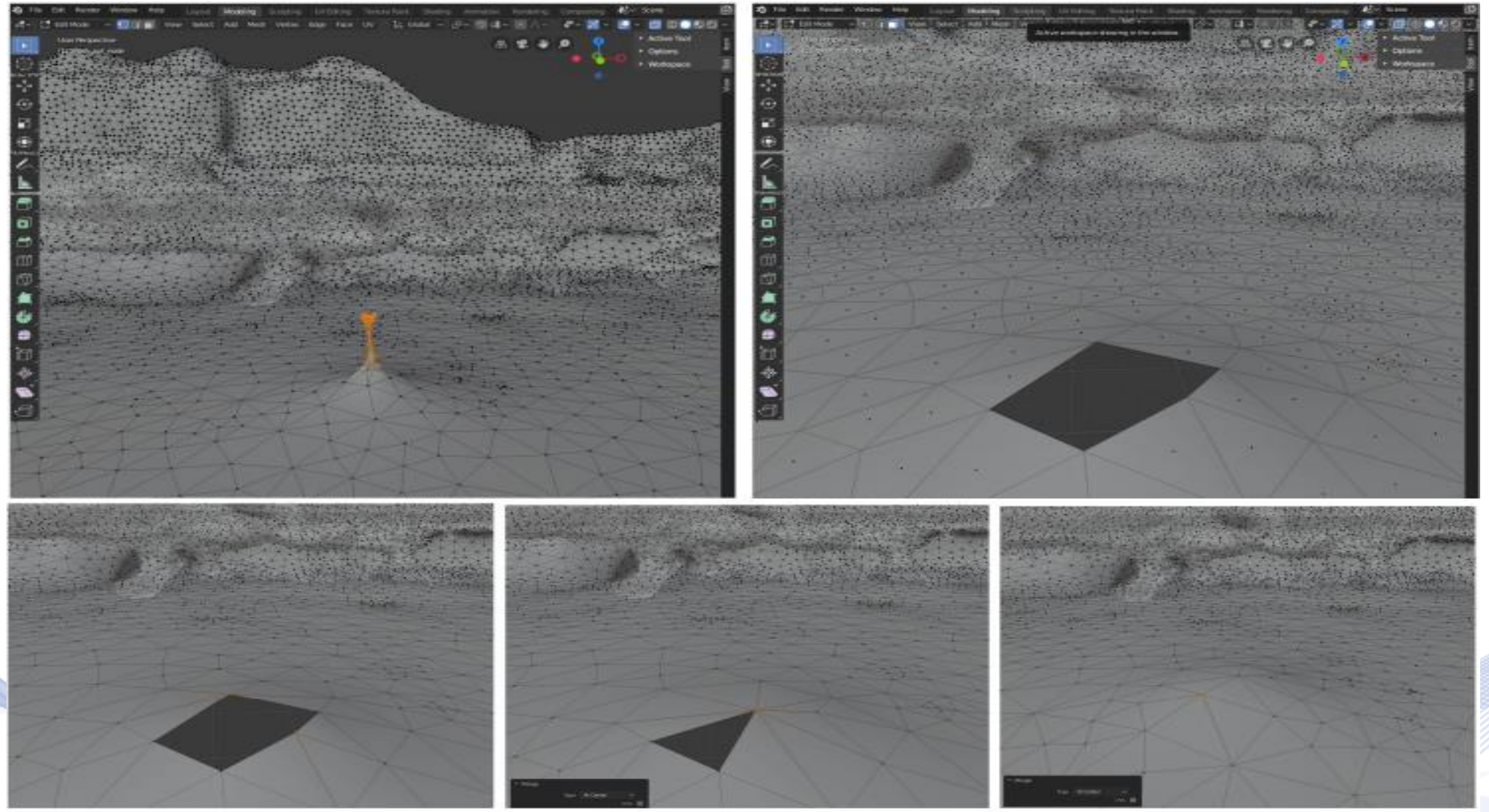


Figure: An example of Smoothing. (Left) The initial 3D model structure. (Right) The processed 3D model structure after Smoothing.

Mesh Modification



Texture Paint

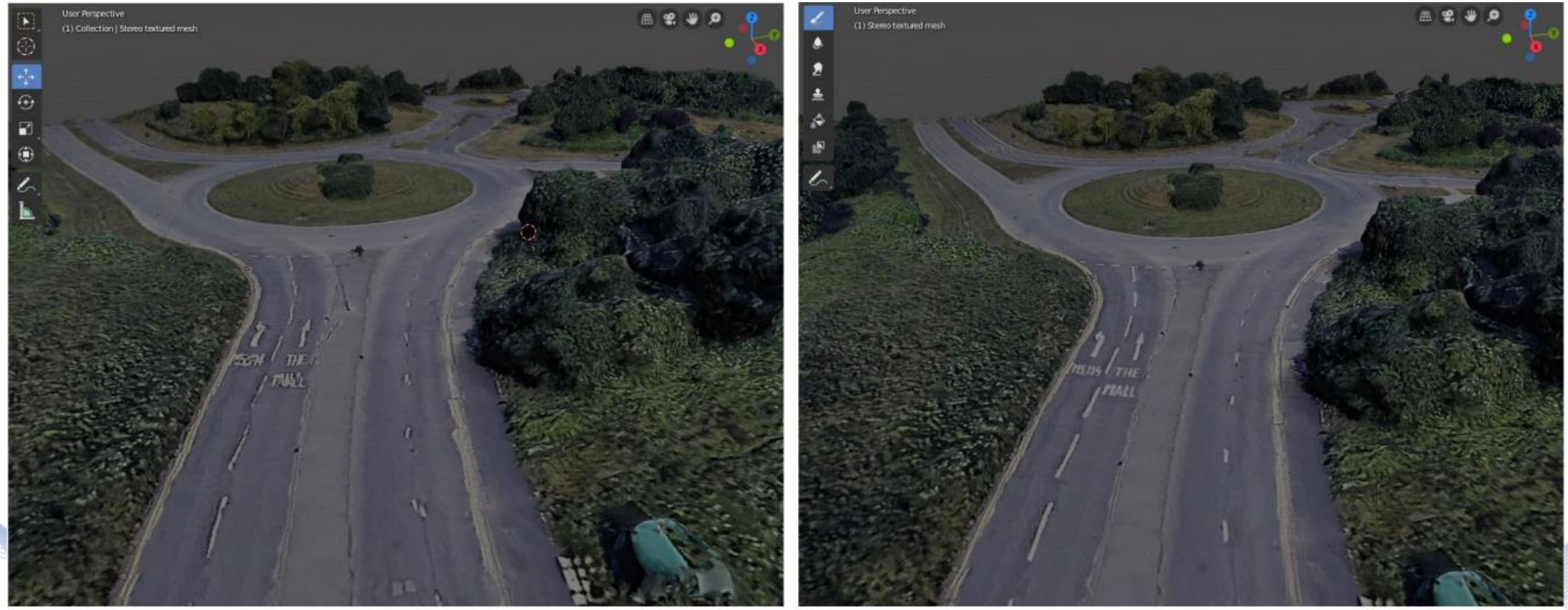


Figure: (Left) A screen shot of the 3D Model before applying Texture Paint. (Right) A screen shot of the 3D Model after applying Texture Paint.

Bibliography

[PIT2021] I. Pitas, “Computer vision”, Createspace/Amazon, in press.
[SZE2011] R.Szelinski, “ Computer Vision ” , Springer 2011
[HAR2003] Hartley R, Zisserman A. , “ Multiple view geometry in computer vision” . Cambridge university press; 2003.
[DAV2017] Davies, E. Roy. “Computer vision: principles, algorithms, applications, learning ”. Academic Press, 2017.
[HEL2002] Copyright Hellenic Ministry of Culture and Sports (L. 3028/2002).
[ARB2013] 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.

[BAY2008] H. Bay, A. Ess, T. Tuytelaars, L. Gool, “Speeded-Up Robust Features (SURF)”, Computer Vision and Image Understanding, 110(3), 346-359, 2008.

[BOI1984] J. D. Boissonnat, “Geometric Structures for Three-dimensional Shape Representation”, ACM Transactions on Graphics (TOG), pp. 266-286, 1984.

Bibliography

[CAL2009] 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.

[CAS1998] 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.

[FUR2015] Y. Furukawa, C. Hernández, “Multi-View Stereo: A Tutorial”, Foundations and Trends® in Computer Graphics and Vision, 9(12), 1-148, 2015.

[LOW2004] G. D. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, International Journal of Computer Vision, 60(2), 91-110, 2004.

[SAL2004] J. Salvi, J. Pagès, J. Batlle ,”Pattern Codification Strategies in Structured Light Systems”, Pattern Recognition, 37(4), 827-849, 2004.

Bibliography

[TRI2020] B. Triggs, B. McLauchlan, R. I. Hartley, A.W. Fitzgibbon, “Bundle Adjustment — A Modern Synthesis”, Berlin, Germany: Springer-Verlag, 2020.

[VRU2009] 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.

[SMI2018] 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.

[SEI2019] 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.

[HAW2016] 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.

Bibliography

[TSC2005] D. 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.

[WON2006] A. Wong and J. Orchard, “A nonlocal-means approach to exemplar-based inpainting”, IEEE International Conference on Image Processing, pages 2600–2603, 2006.

[BLA] 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
[SAN] A. M, 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

Q & A

Thank you very much for your attention!

More material in
http://icarus.csd.auth.gr/cvml-web-lecture-series/

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