Photogrammetry and Machine Learning for Cultural Heritage

Dr. Panagiotis Agrafiotis
pagraf@central.ntua.gr
http://users.ntua.gr/pagraf

Postdoctoral Researcher
Lab. of Photogrammetry, Signal Processing & Computer Vision
School of Rural and Surveying Engineering
National Technical University of Athens, Greece

STABLE Summer School 14-18/9/2020, Rethymno, Crete, Greece

How we can **get geometric** and **semantic information** from **images**?

What is Photogrammetry (and 3D Computer Vision)?

Definitions in the literature:

- "**Photogrammetry** is the art and science of determining the position and shape of objects from photographs[1]"
- "**3D Computer Vision** is a mathematical technique for recovering the 3D shape and appearance of objects in imagery[2]"

- Same thing, different equations - Photogrammetric computer vision

[1] K. Kraus, Photogrammetry, 1, Dümmler (1994)
[2] R. Szeliski, **Computer vision: algorithms and applications** Springer Science & Business Media (2010)

Why we need it?

- **Cheap**, **fast** and **extremely precise***:
  - Results in **extremely dense and precise 3D surface data**
    - limited number of photos
    - standard digital photography equipment
    - In a short period of time.
- **3D supports many activities**:
  - Small/large **object** or small/large **area 3D** documentation,
  - Statistical analysis
  - Historical reconstructions
  - **Communication** and **promotion** of the sites
  - Accessibility to remote sites (i.e. underwater CH comes to reach through VR)

*Results need to be controlled and critically considered to be reliable. It must be applied by experts.

What has to offer in structural stability of historical buildings?

- **Non-invasive**
- **Delivers complete, very detailed and very accurate 3D models**

  - Feed these models into **stability control monitoring models** (i.e. finite element model (FEM) analysis)

  - **Overlay additional information** and data on the models (thermal hyperspectral and imagery, GPR measurements etc.)

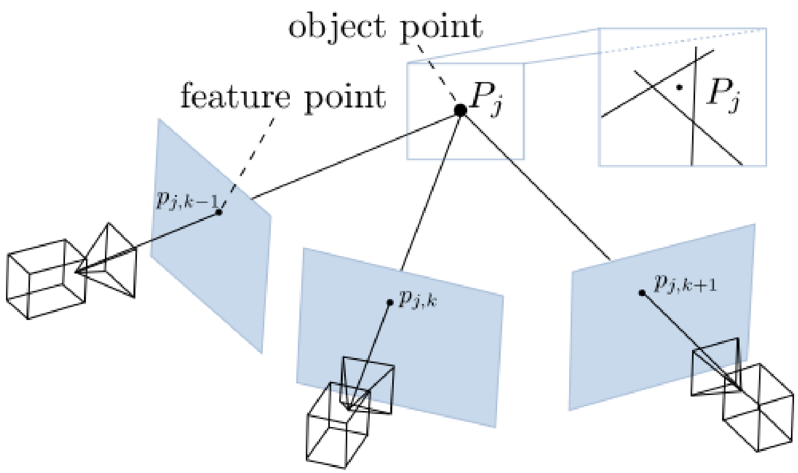
  - **Use the visual and semantic information** for generating sections and drawings of different levels of generalization and detail

Thus, it is essential part of the Cultural Heritage documentation processes <u>overwater and underwater</u>.

How does it work?

- **It is based on the collinearity condition**: Bundles of light rays pass from object points through the perspective center of the camera lens onto their corresponding image points.
- Using two images from different camera positions, 3D object coordinates can be determined by intersecting the reconstructed bundles of light.

- **Nowadays** it is based on Structure-from-Motion (SfM) and Multi-View Stereo Techniques (MVS)



Source: left: imagine.enpc.fr/~moulonp/openMVG/Triangulation_n_geometry.png right: https://vision.in.tum.de/research/image-based_3d_reconstruction/multiviewreconstruction

The problem

- 2D images are the projection of the 3D space into a 2D surface: **3D information is missing**
- 3D reconstruction of the space is the opposite process: **3D space reproduction after a set of overlapping 2D images (depth information retrieval)**

- **Solution**:

Image matching

Seeking to match points between the images

\+

Triangulation

Calculating 3D coordinates based on the successful matches

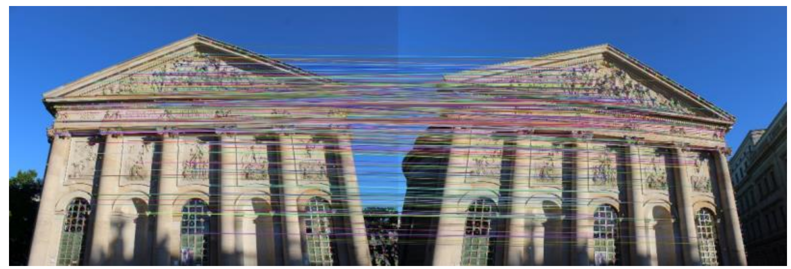
Basic procedure

Data collection

- Images (>3)
- Ground Control Points (GCPs)

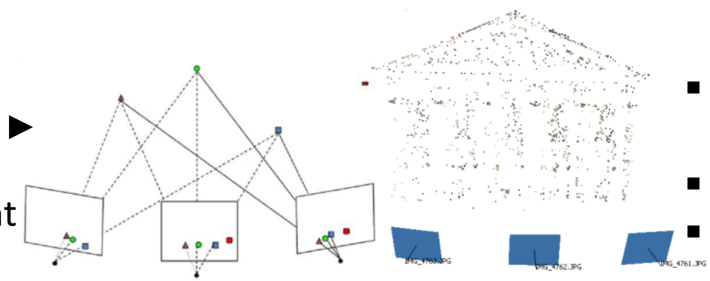
Structure from Motion

- Key-point detection
- Key-point matching and filtering



- 2D matches

- Image orientation
- Triangulation
- Bundle adjustment



- Sparse 3D point cloud
- Camera positions
- Camera calibration

Multi-View Stereo

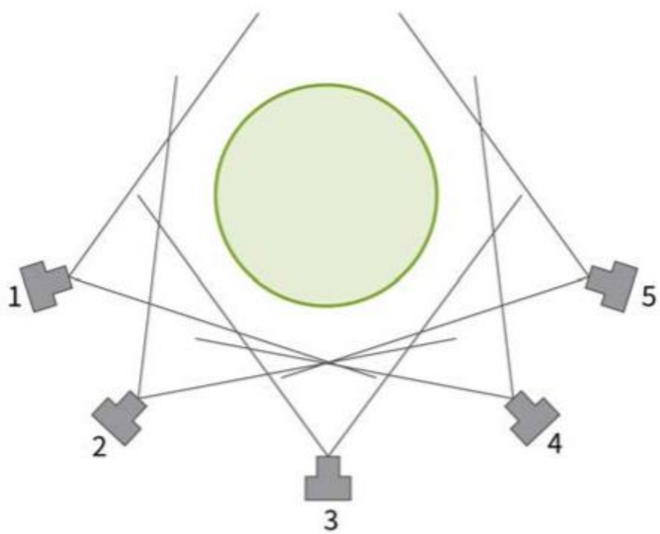
- Dense image matching

- Disparity maps
- Dense 3D point cloud

3D reconstruction steps in depth…

Image collection

- Overlapping images with appropriate geometry
- Extra images to strengthen the geometry of the block
- HOWEVER, not too many images since this will increase the computational complexity of the process
- Also important the equipment (camera etc.)



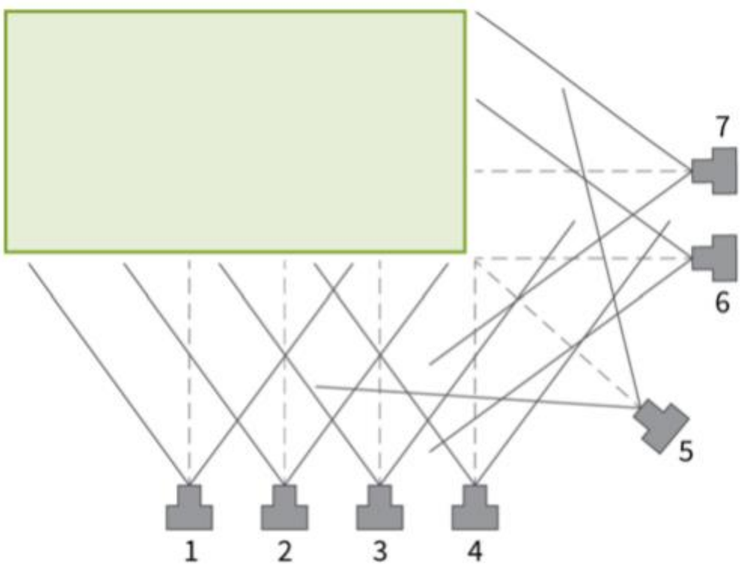


Image Source: https://historicengland.org.uk/images-books/publications/photogrammetric-applications-for-cultural-heritage/heag066-photogrammetric-applications-cultural-heritage/

Key-point detection & description [1]

Feature-based

- Detection: Moravec, Förstner, Harris κ.α
- Description: Brief, Daisy, Brisk, Freak κ.α.
- Most famous packages: SIFT & SURF

Phase based

- Used mostly in aerial photogrammetry
- They are facing difficulties when dealing with large changes of viewpoints and perspective distortions

Key-point detection & description [2]

Key-points need to be

- Well localized
- Distinctive
- In large amounts

- Geometric invariant: translation, rotation, scale
- Photometric invariant: brightness, exposure, …

Image Source: Algorithms and Applications Richard Szeliski, 2010

Key-point detection & description [3]

Example of areas for distinctive and non-distinctive features

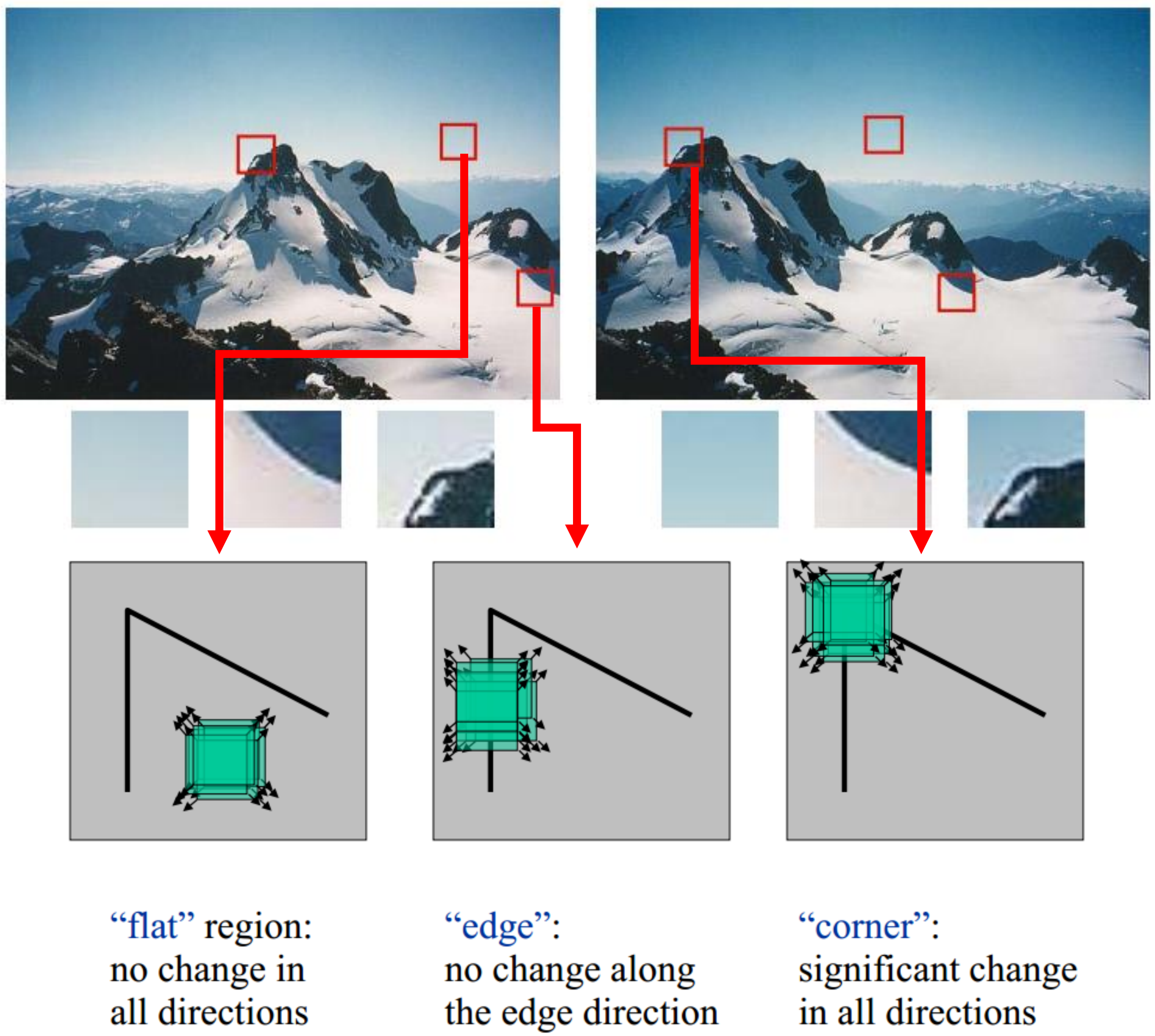


Image Source: Algorithms and Applications Richard Szeliski, 2010

Detected key-points

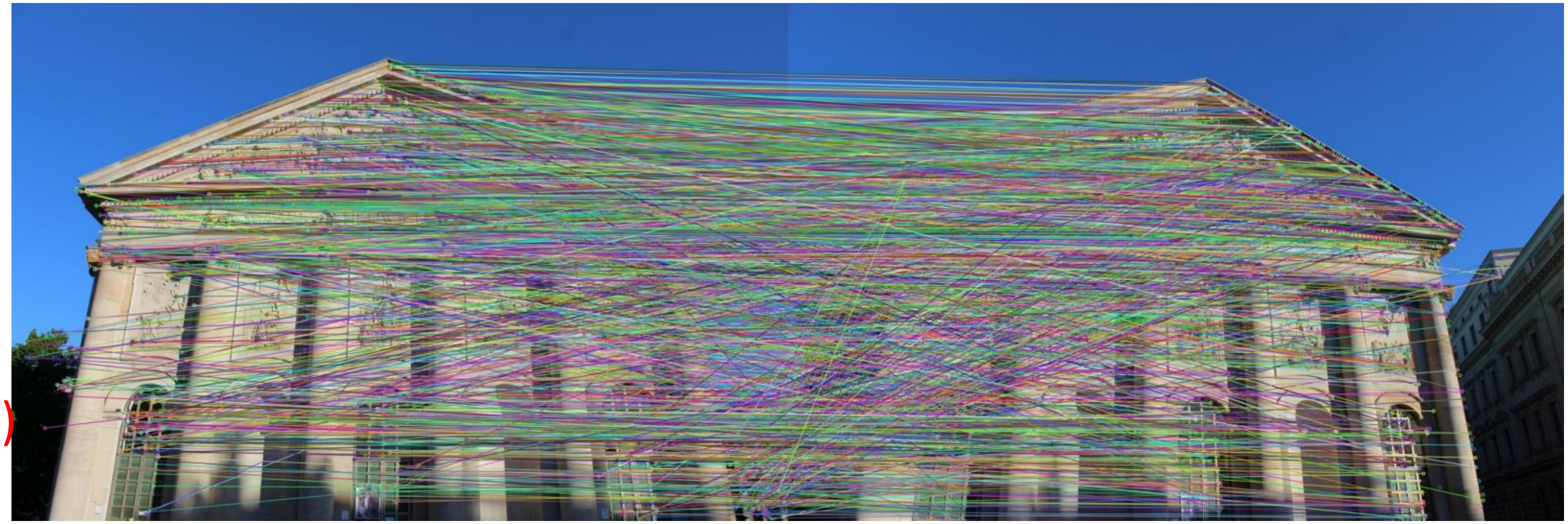
Missing from areas
without texture

Key-points matching

- Brute-Force matching
  - It takes the descriptor of a point and calculates the matching distance with every point in the other image. The closest one is returned.

- FLANN matching (Fast Approximate Nearest Neighbor Search Library)
  - May be faster but less accurate but much faster than Brute-Force

Unfiltered (Brute-Force) Lines are intersecting!!!)



Filtering Matches

- **Distance ratio test** to try to eliminate false matches (usually set to 0.8)
- **Cross check test**
- **Geometric test** (eliminate matches that do not fit to a geometric model, e.g. RANSAC or robust homography for planar objects)
- **<u>Usually all the above are used together</u>**

Filtered (all the above)

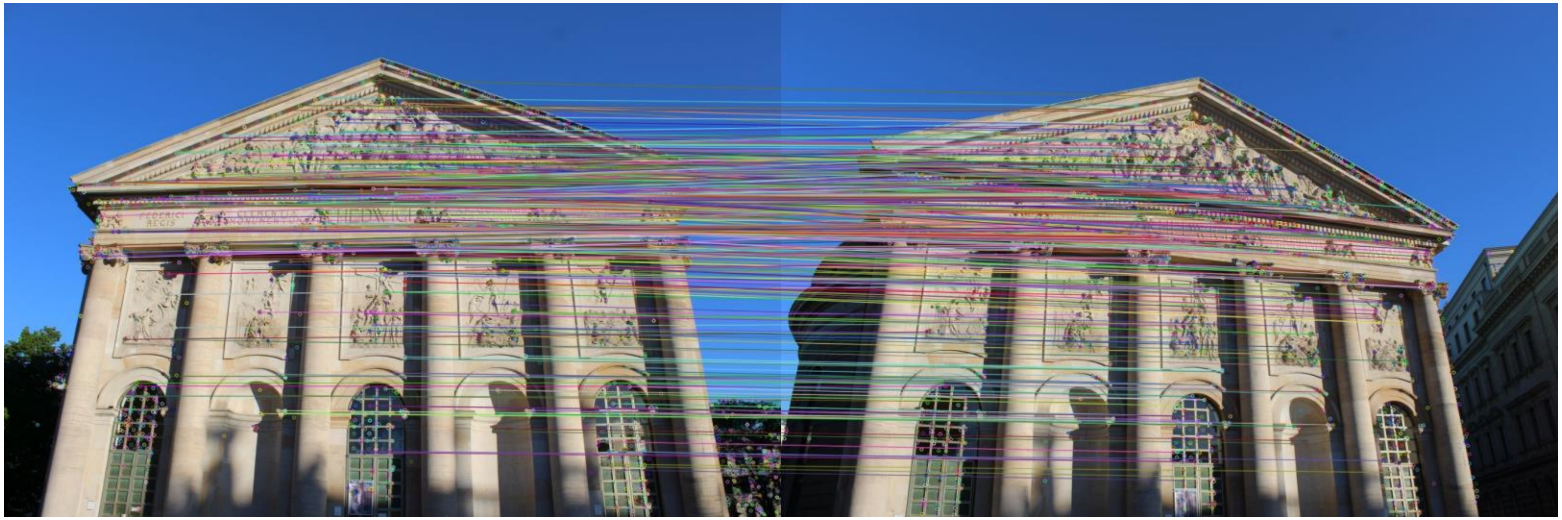
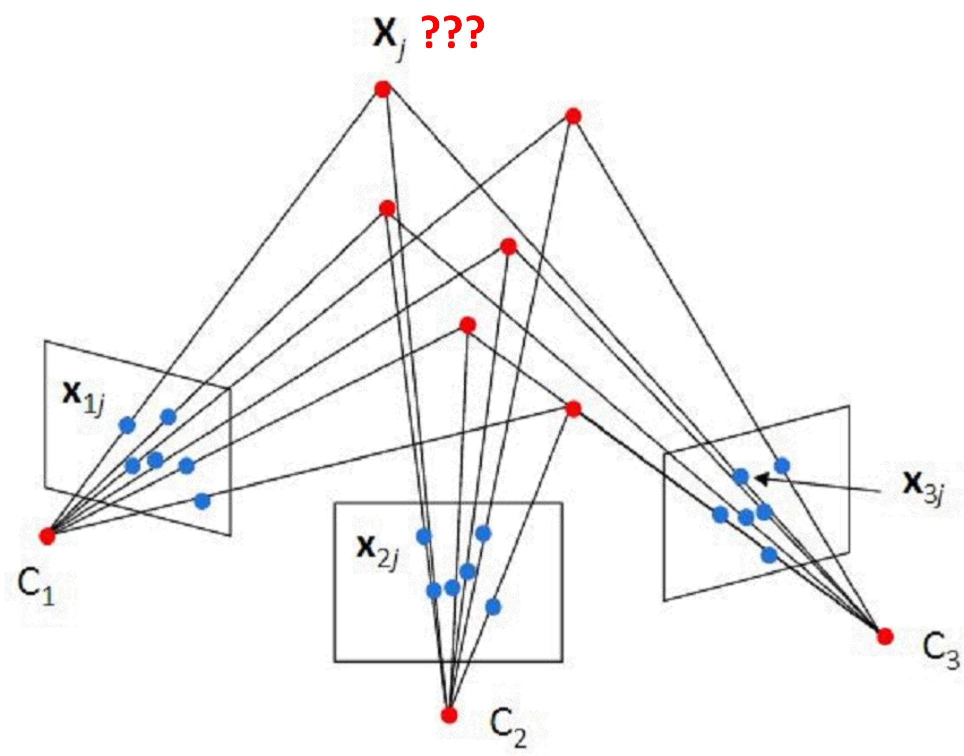


Image orientation, Triangulation and bundle adjustment

- The matched points are forming the **sparse point cloud**
- **Camera positions** Xo, Yo, Zo
- Camera (self) **calibration**

- If GCPs have been used and measured on the images, the camera positions and the sparse point cloud are in a geodetic reference system

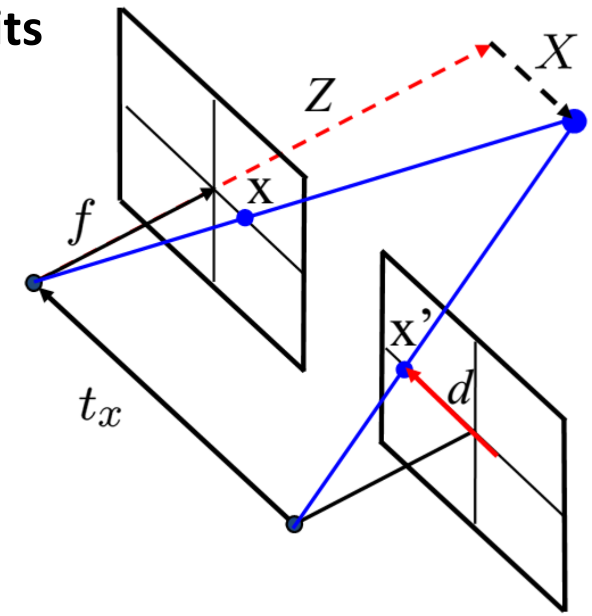


Depth (disparity) maps

What is disparity?

Disparity ***d*** is the distance between a pixel and **its horizontal match** in the other image.

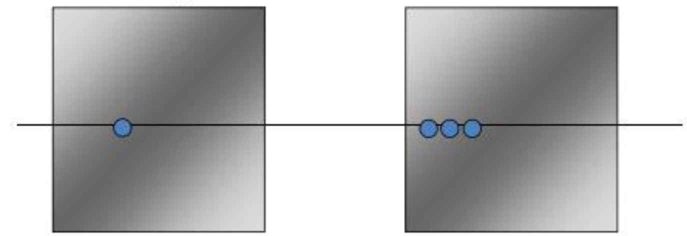
- To get disparities, **dense image matching methods are used** (Semi-Global matching, window-based matching etc.)
- From each stereopair, one disparity map is generated
- When merged, they can give the final 3D dense point cloud



Problems in Depth (disparity) maps generation

- Camera issues
  - Noise
  - Lens distortions
  - Chromatic aberration
- Camera position and geometry issues
  - Perspective distortions
  - Occlusions
  - Specularities
- Issues related to the scene
  - Changes in illumination
  - Moving objects

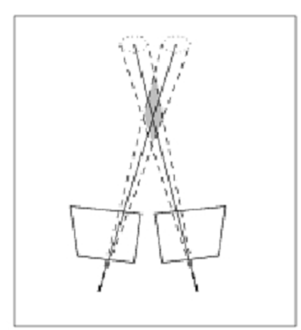
- Still, matching can be ambiguous!
  - Low textured regions



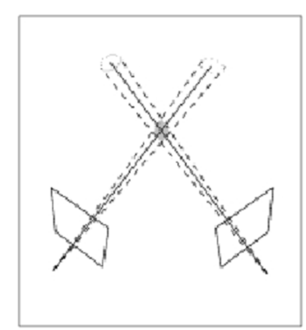
  - Repetitive texture pattern

Geometry issues

Easy matching depth uncertainty



Difficult matching, less depth uncertainty



Dense 3D point cloud

(Dense 3D point cloud from just 3 images…)

In general, color is obtained by the pixels matched to generate the point (average etc.)

Notes!

- **3D dense point clouds can be then used for 3D triangulation, 3D mesh generation and texturing**

- The accuracy of the 3D reconstruction is subject of a variety of factors:
  - Image quality
  - Overlapping and image base
  - Ground Control Points’ accuracy in measurement and marking on the images
  - The quality and reliability of the algorithms used
  - The level of expertise of the user

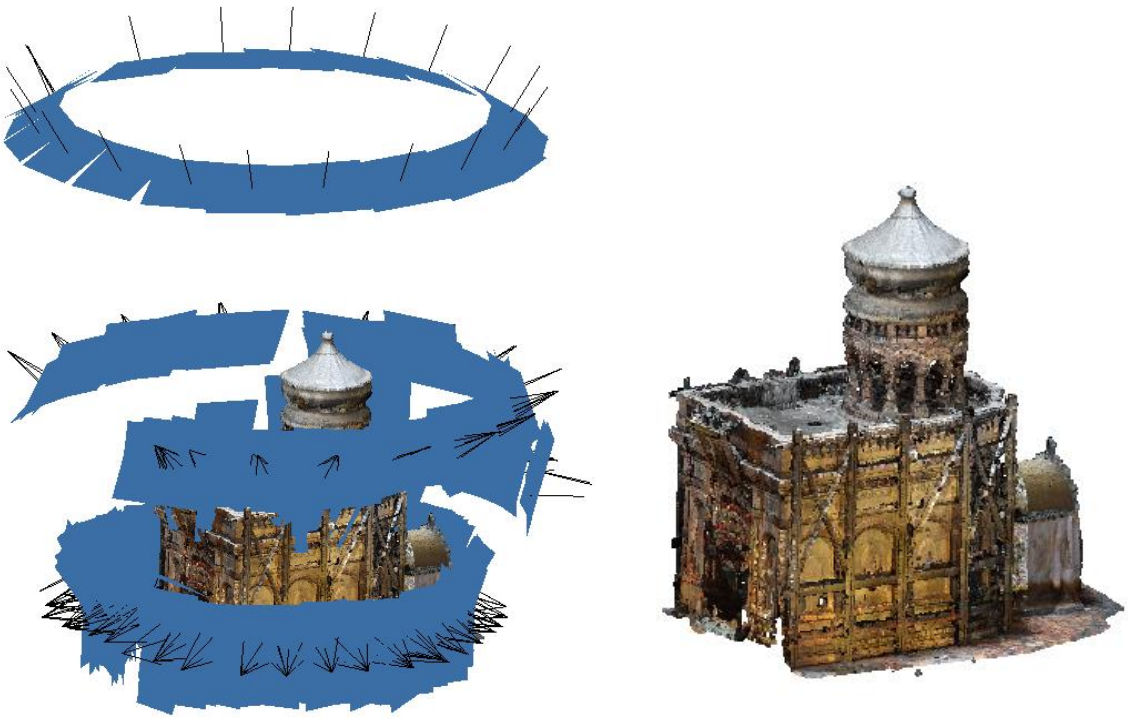
Sources of errors in the process

- **Systematic errors**
  - Are related to sensor and camera geometry
    - Non-planar sensors
    - Non-square pixels
    - Non-perpendicular lens axis and sensor
    - Incorrect camera calibration

- **Random errors and mistakes**
  - Wrong image alignment
  - Wrong GCPs’ measurements and marking

Building Examples [1]

- 3D reconstruction of the Holy Aedicule in Jerusalem for restoration works
- Tomb was at risk of ‘catastrophic’ collapse



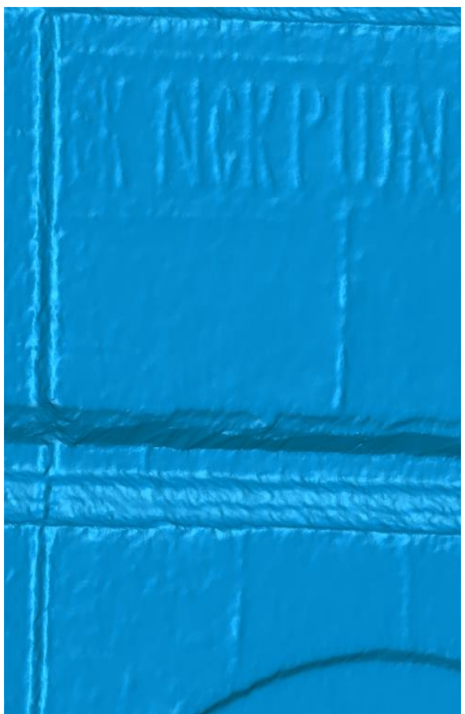
- Necessary the fast 3D documentation
- Interdisciplinary collaboration
- Continuous monitoring during the restoration

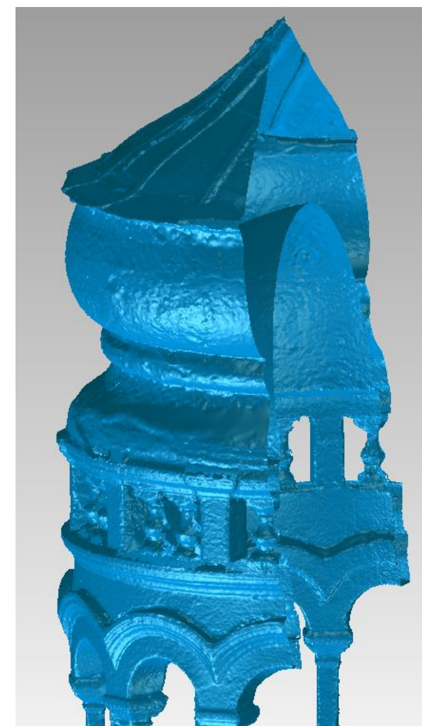
Building Examples [1] - difficulties

- Obstacles: Visitors, scaffoldings etc
- Really narrow spaces – use of wide-angle lens
- Smoke and different illumination spots from candles etc.
- Restrictions on lighting – capturing during night due to the pilgrims

Building Examples [1] - Details

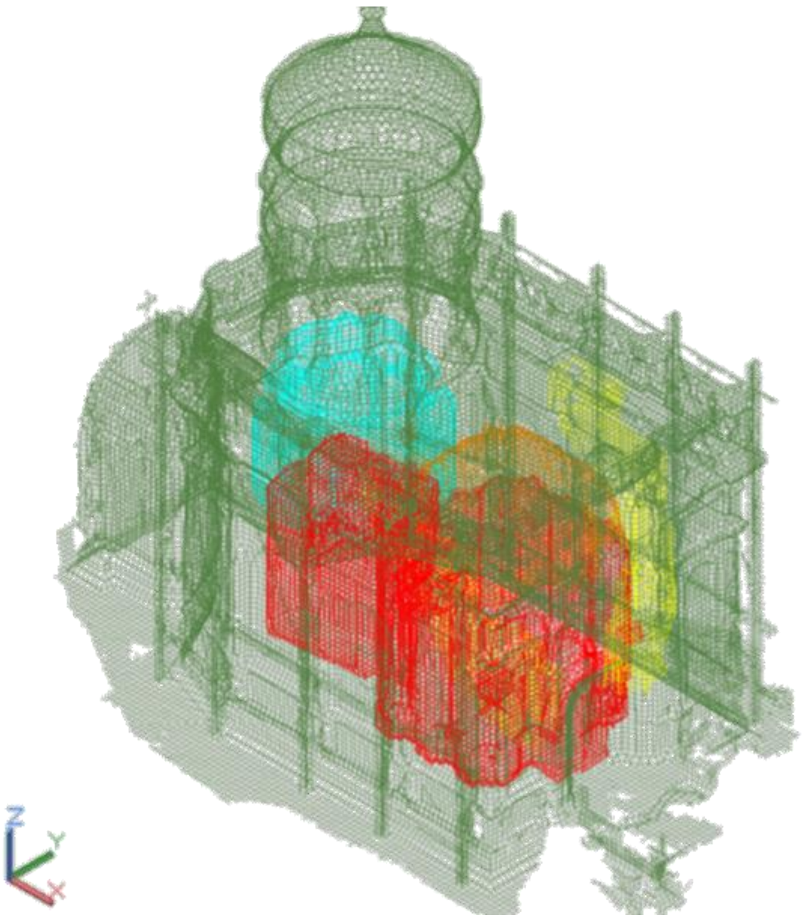
- Very high accuracy and detail on the 3D reconstruction
- Building deformations were identified
- Damages were identified

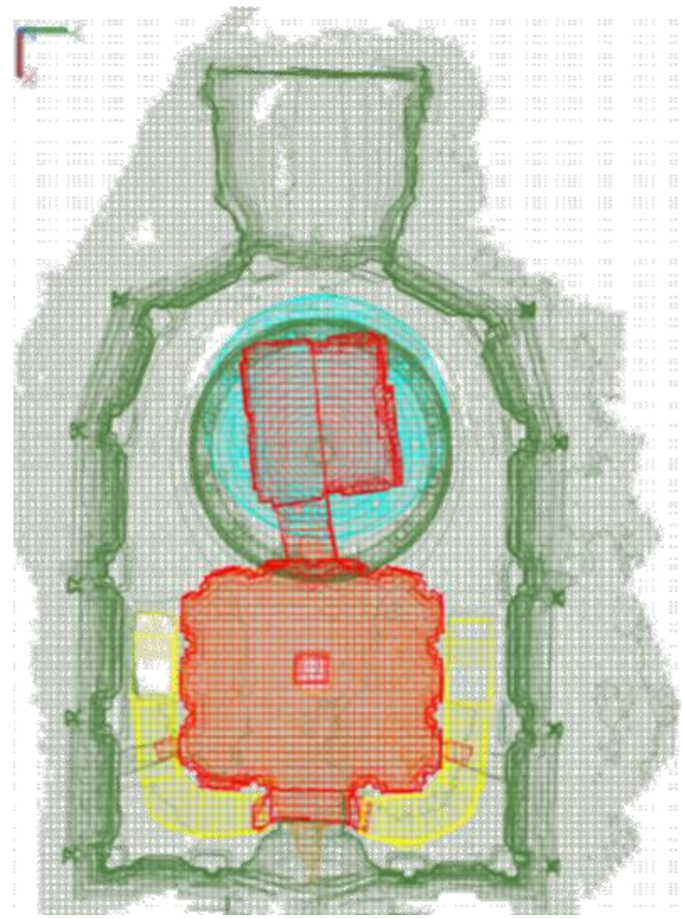


Building Examples [1] - FEM

- Sections per 10cm for the Finite Element Analysis

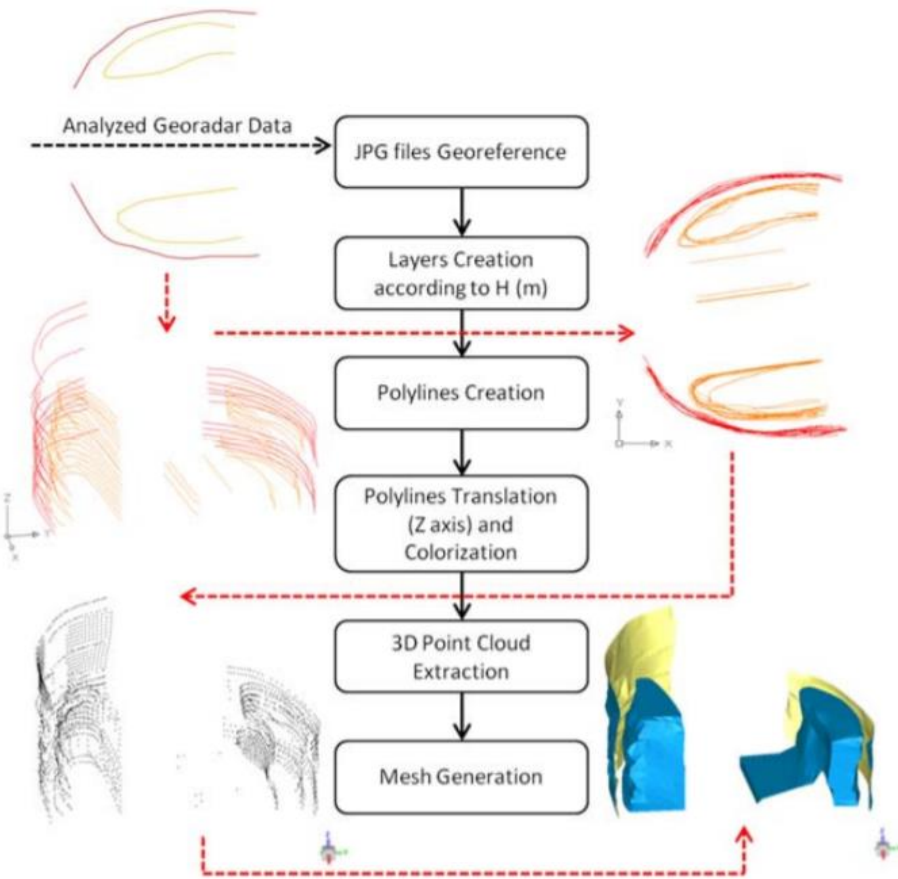




Building Examples [1] - restored

Building Examples [2]

- Fusion of the 3D model with GPR measurements in order to create the **3D of the interior of the walls (Holy rock)**, to feed the Finite Element Models



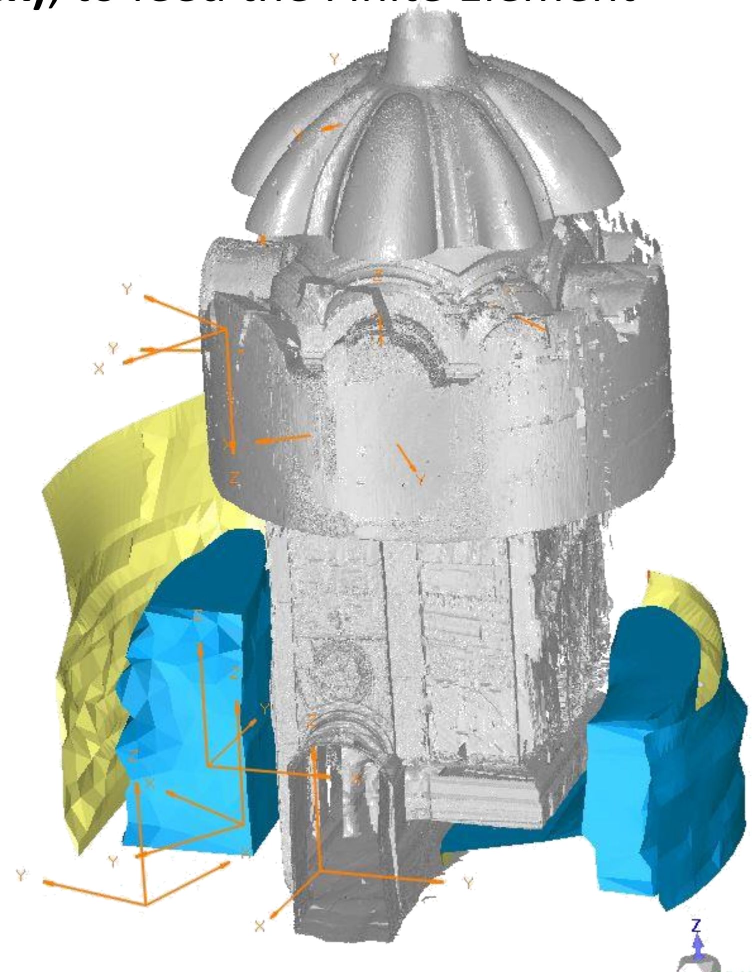


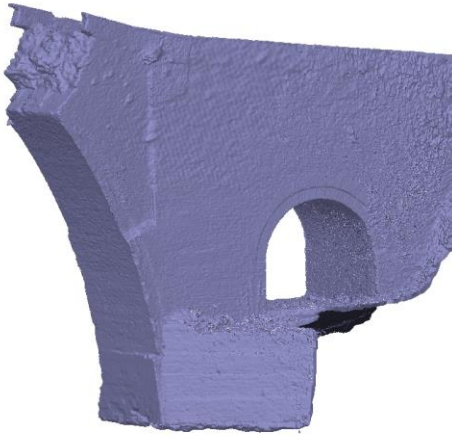
Image Source: Agrafiotis, P., Lampropoulos, K., Georgopoulos, A., and Moropoulou, A.: 3D MODELLING THE INVISIBLE USING GROUND PENETRATING RADAR, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W3, 33–37, https://doi.org/10.5194/isprs-archives-XLII-2-W3-33-2017, 2017.

Building Examples [3] – the bridge

- Plaka bridge is the largest one-arc bridge of Balkans
- Partially destroyed during a flood, when a huge tree hit on it

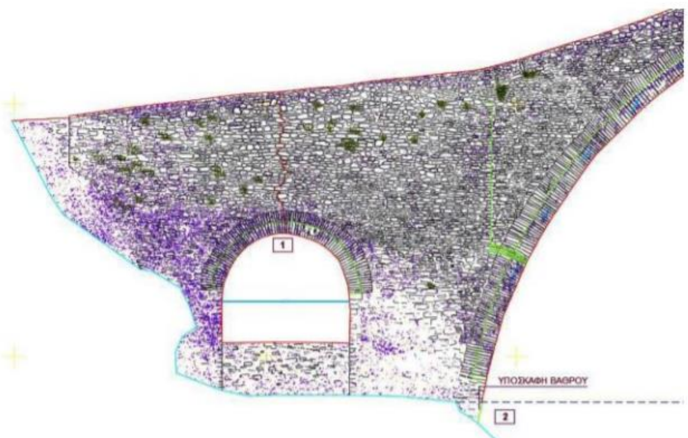
Building Examples [3] – 3D

- 3D reconstruction of Plaka bridge to facilitate it's restoration

Building Examples [3] - products

- 2D Drawings
- Pathologies plans
- Retrieval of the main arch characteristics



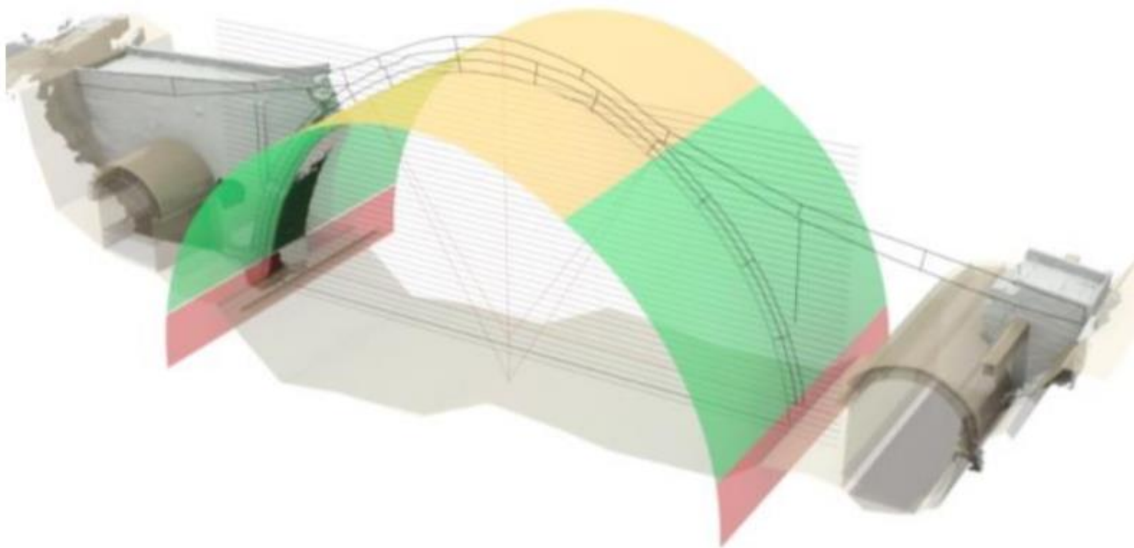


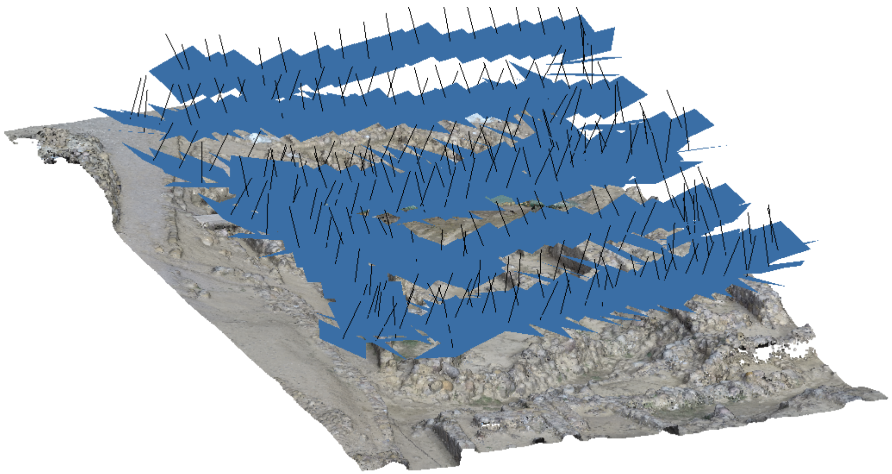
Image Source: Kouimtzoglou, T., Stathopoulou, E. K., Agrafiotis, P., & Georgopoulos, A. (2017). Image-based 3d reconstruction data as an analysis and documentation tool for architects: the case of Plaka Bridge in Greece. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 42, 391.

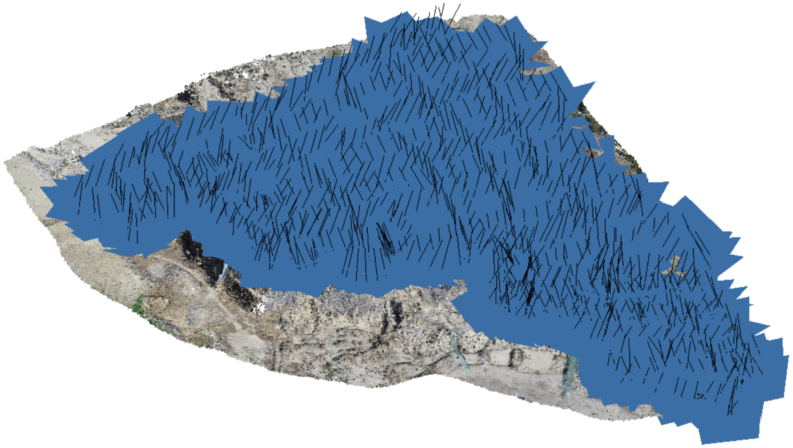
Building Examples [3] - restored

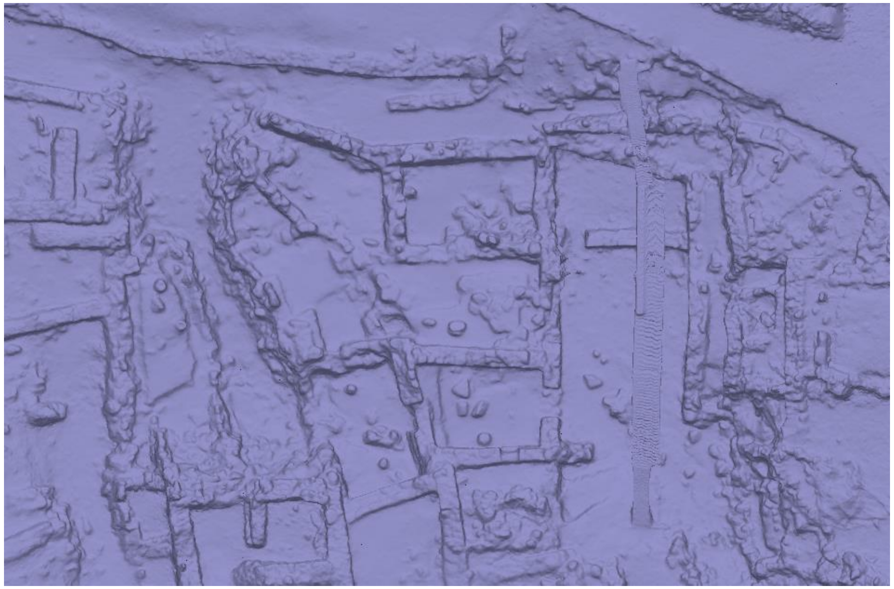
Image Source: https://www.amna.gr/en

Other Examples [1]

- 3D reconstruction of small or large archaeological sites

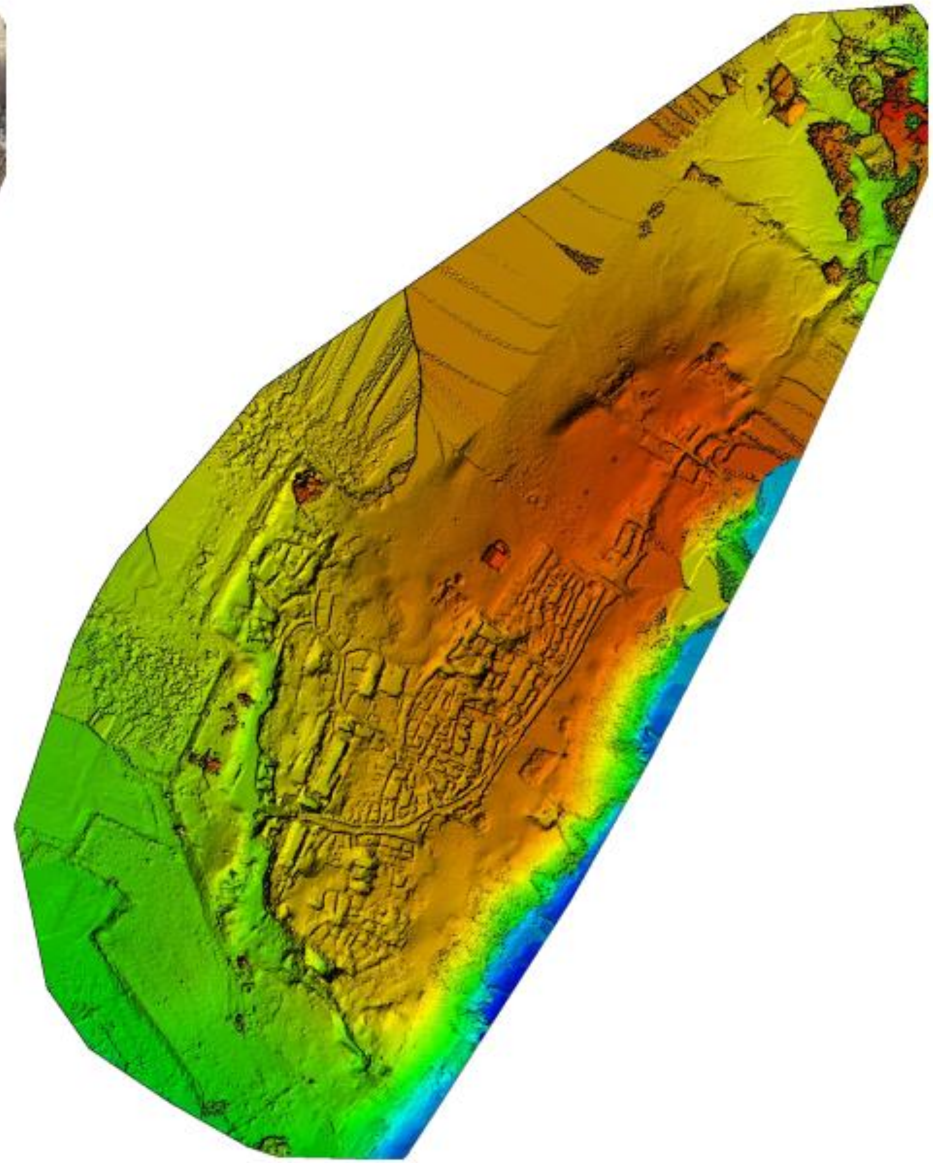






Other Examples [2]

- Orthoimages and DEMs

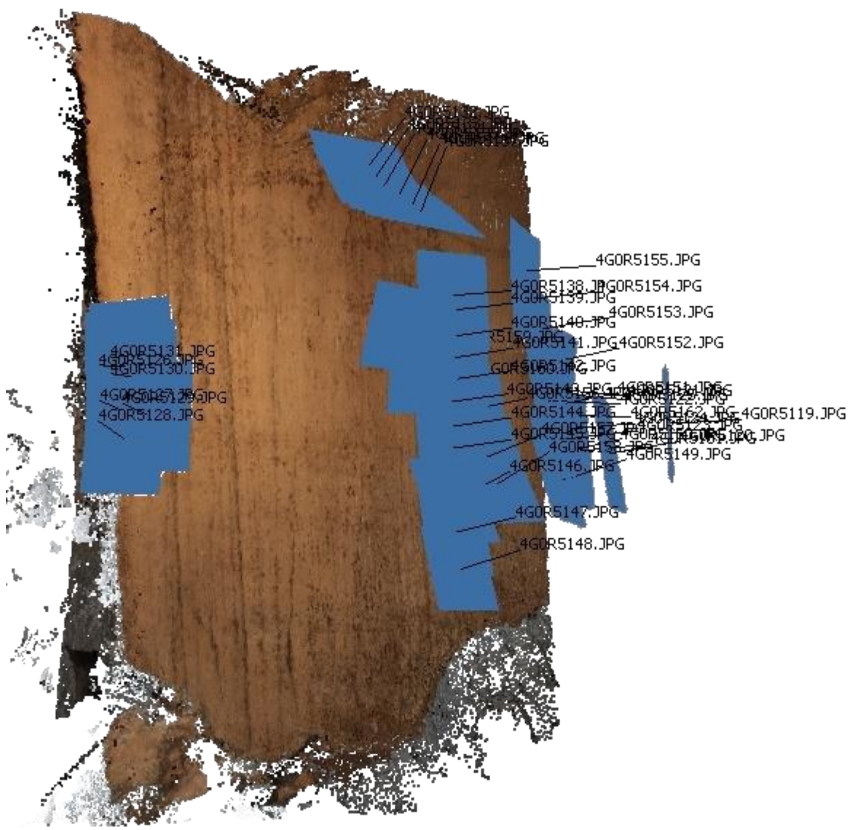


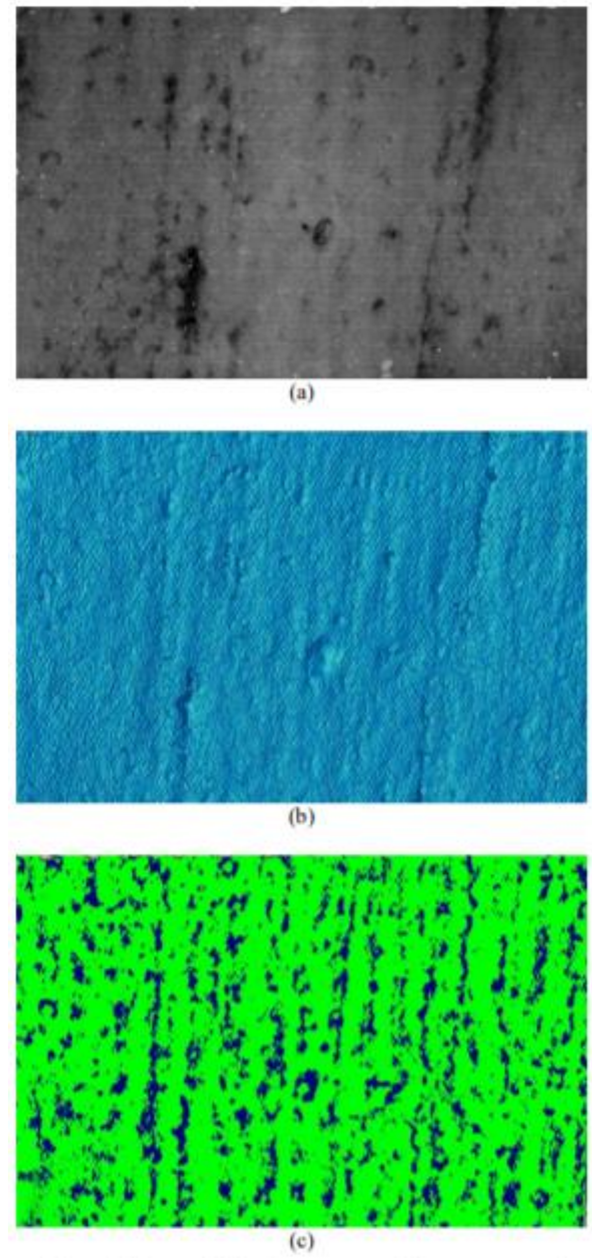
Other Examples [3]

- 2D plans and/or orthoimages

Other Examples [4]

- 3D reconstruction of Parthenon inscriptions
- Detect weathered letters in areas that wasn't known that contained letters





Source: Papadaki, A. I., Agrafiotis, P., Georgopoulos, A., & Prignitz, S. (2015, February). Accurate 3D scanning of damaged ancient Greek inscriptions for revealing weathered letters. In International Workshop on 3D Virtual Reconstruction and Visualization of Complex Architectures, 3D-ARCH 2015.

Underwater photogrammetry

- Mapping submerged CH using underwater imagery

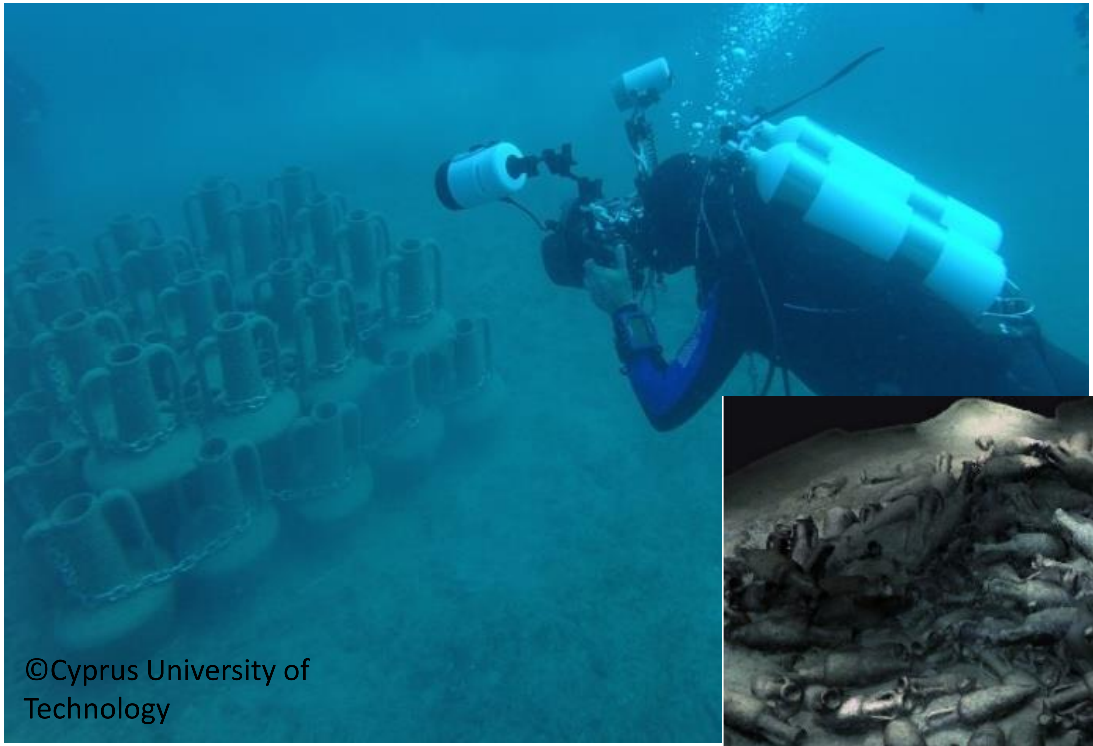
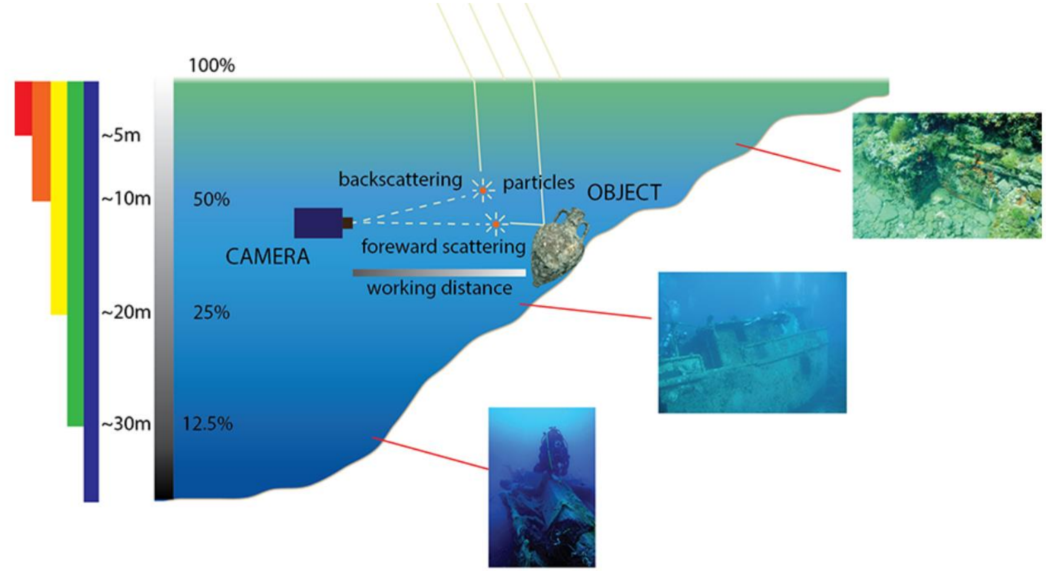


Image Source: down right: Drap, P., Seinturier, J., Hijazi, B., Merad, D., Boi, J. M., Chemisky, B., ... & Long, L. (2015). The ROV 3D Project: Deep-sea underwater survey using photogrammetry: Applications for underwater archaeology. Journal on Computing and Cultural Heritage (JOCCH), 8(4),

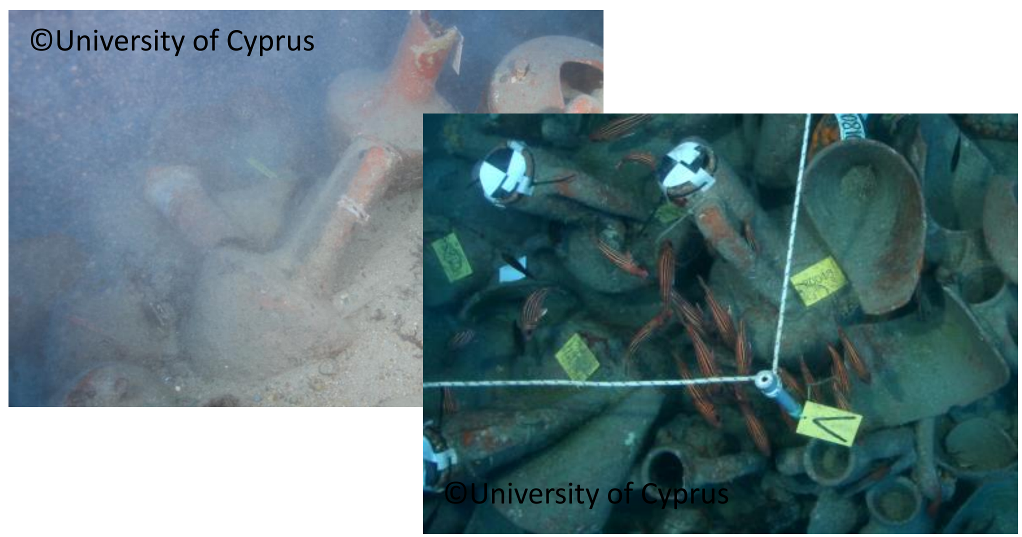
Underwater photogrammetry - difficulties

- Refraction
- Colors are lost
- Chromatic aberration
- Visibility
- Turbidity
- Sea life
- Limited time

- Equipment
- Georeferencing and scaling



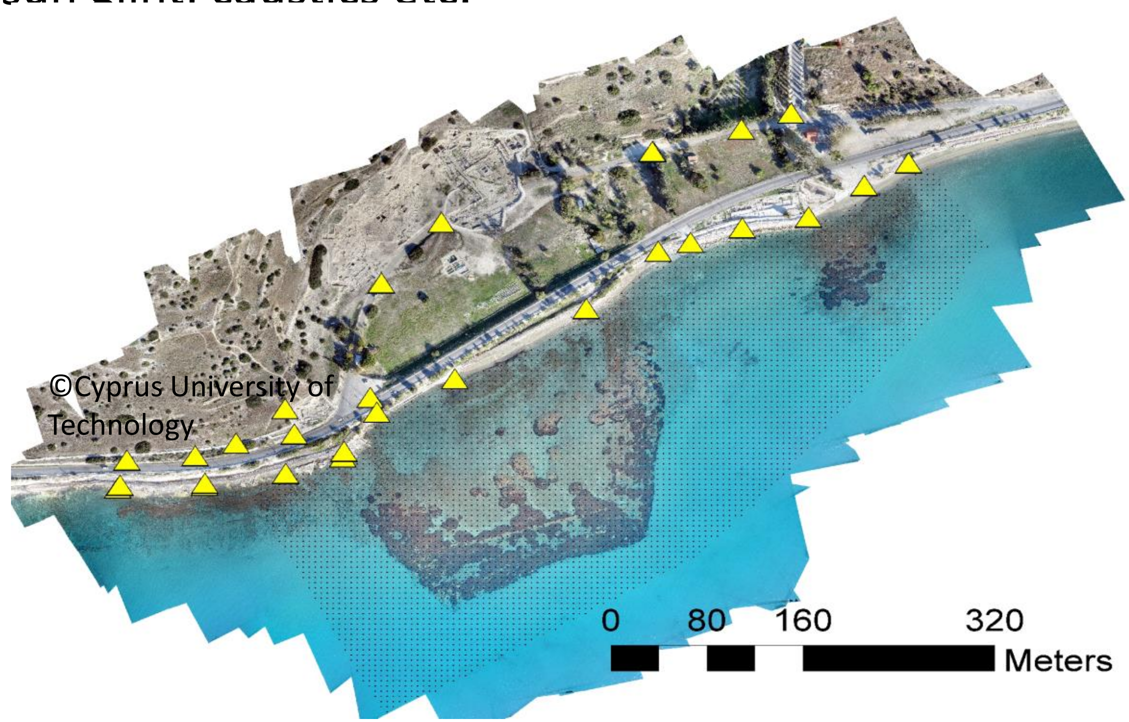
Source: G. Bianco , M. Muzzupappa , F. Bruno , R. Garcia , L. Neumann , A NEW COLOR CORRECTION METHOD FOR UNDERWATER IMAGING



Overwater photogrammetry for shallow underwater CH areas

The case of an ancient submerged port

- Mapping submerged CH using overwater imagery
- **<u>Major problem the refraction effect – delivers erroneous depths</u>**
- Waves, sun glint, caustics etc.

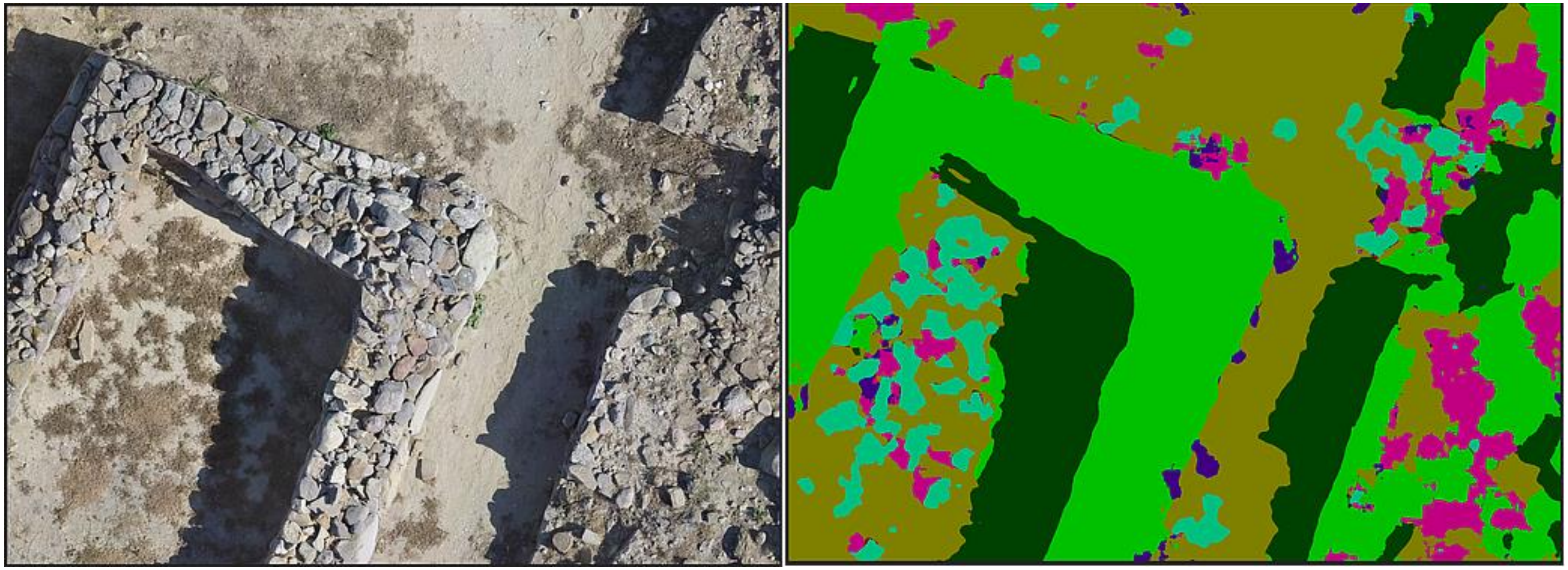


What’s next?

What else we can get using these outputs ?

Further needs

- Need for automated and fast **semantic information** on the 3D point clouds, meshes (to be used in BIM) and orthoimages.
- Need to use this **semantic information** in photogrammetric processing to increase the accuracy.



Machine Learning for CH

- **Machine and Deep Learning offer the tools to satisfy these needs**.

However the use of ML in CH **is still limited:**

  - ML is **commonly applied as a ’black-box’ on small datasets that are not generally publicly available**.
  - **Social** and **technical** barriers, strongly related to the **quality** and to the **access of datasets** collected by CH researchers

Solution:

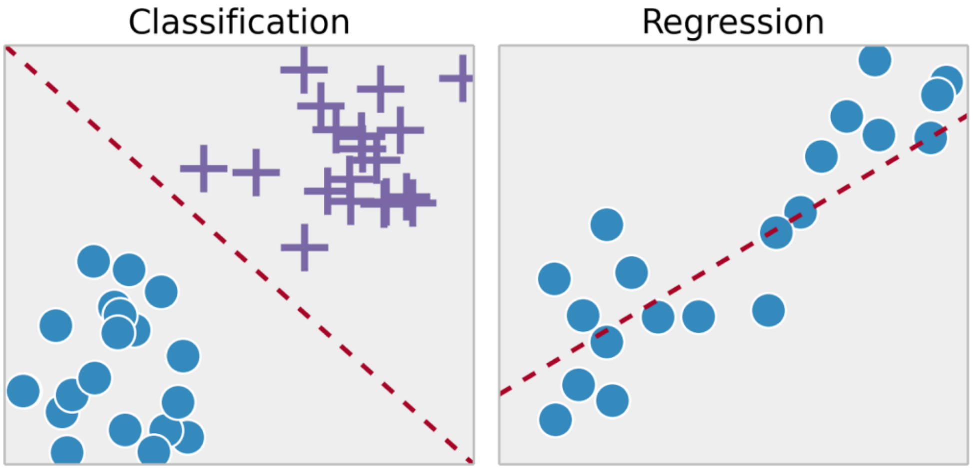
- **Annotated and freely available datasets**
- **Education on ML/DL frameworks**

ML/DL is divided in several categories with the most important the Supervised, the Semi-Supervised and the Unsupervised Learning

Supervised Learning

- **Supervised learning (SL)** aims to learn a function *f* from an input space X to an output space Y given a finite sequence of input-output pairs, called the training set.

- SL algorithms can be divided into regression and classification, based on the nature of the output space.



- **Classification**: get **discrete outputs** of the function *f*
- **Regression**: get **continuous outputs** of the function *f*

Image Source: https://towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d

Most famous SL methods

Shallow architectures

- **Linear and Logistic Regression**
  - Used mainly as a binary classification method
- **Decision Trees and Random Forests**
  - Random Forests are ensemble of different de-correlated decision trees
  - **They are often used in CH** to classify artefacts, archaeological sites and building parts since they are very fast and easy to interpret
- **Support Vector Machines** (also **Support Vector Regression**)

deep architectures

- Supervised **Neural Networks** (CNNs, FCNs etc.)
  - Ability to learn high-level features. Since for CH there is often a lack of large labelled datasets, researchers tackle the feature learning task following a transfer learning approach, where the last layers of a pre-trained network are fine-tuned on the target CH dataset.

Semi-supervised and Unsupervised Learning

- Leverage both labeled and unlabeled data to improve learning performance
  - Semi-supervised Deep Neural Networks

- **Not widely used for CH applications yet…**

Image Source: https://towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d

Examples [1]

- RFs to classify building parts (however, training is performed on part of the same dataset)

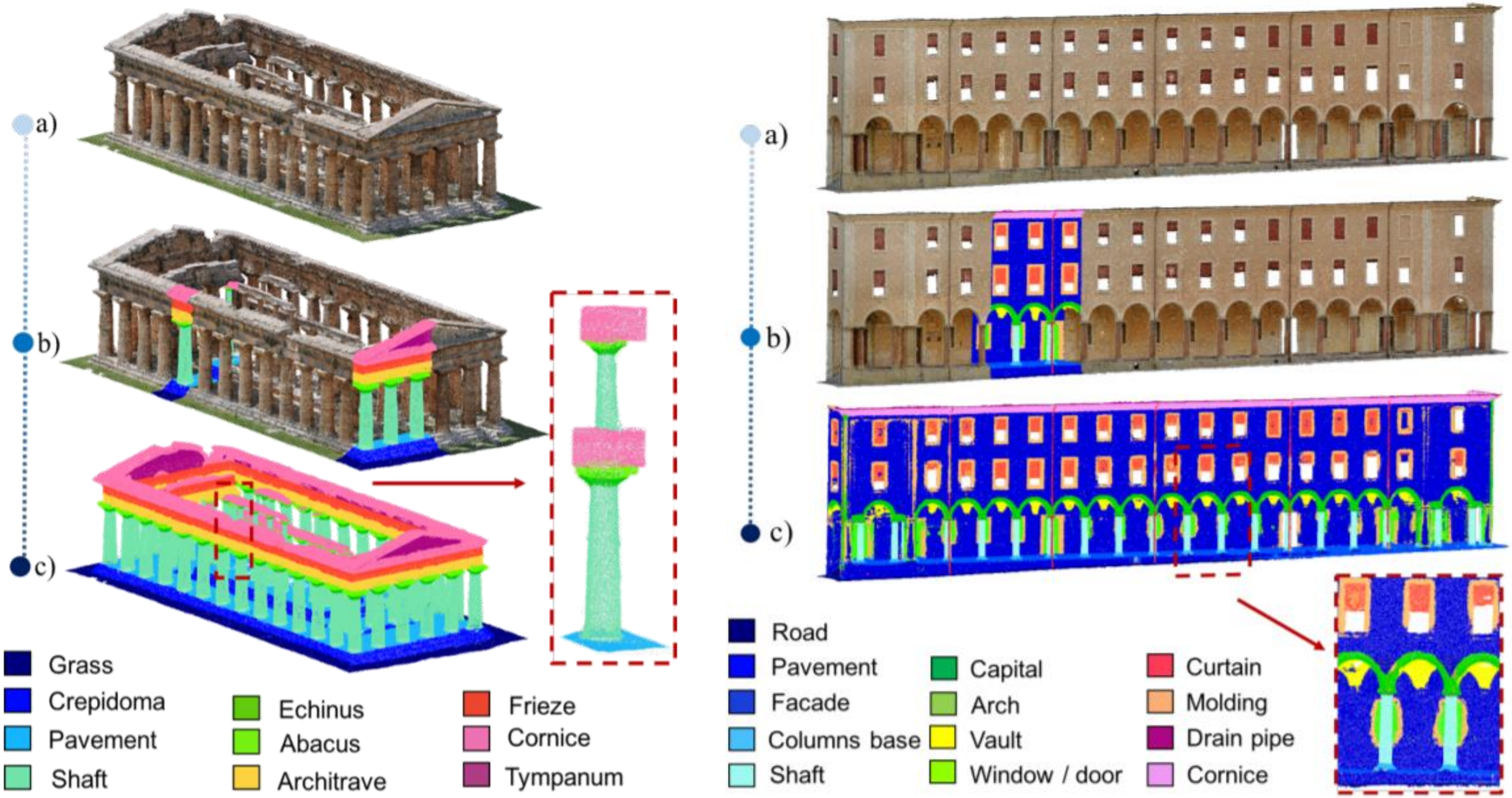


Image Source: Grilli, E., Farella, E. M., Torresani, A., & Remondino, F. (2019). GEOMETRIC FEATURES ANALYSIS FOR THE CLASSIFICATION OF CULTURAL HERITAGE POINT CLOUDS. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences.

Examples [2]

- CNNs to classify building parts and transfer the semantic information on the point cloud

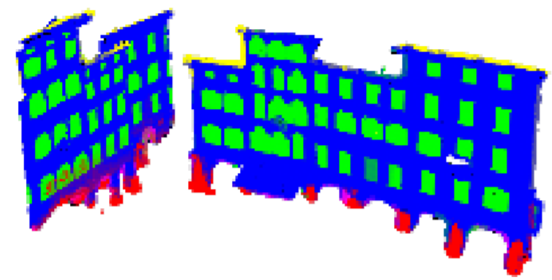


Image Source: Stathopoulou, E. K., & Remondino, F. (2019). Semantic photogrammetry: boosting image-based 3D reconstruction with semantic labeling. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci, 42(2), W9.

Examples [3]

- RFs to classify large buildings' parts

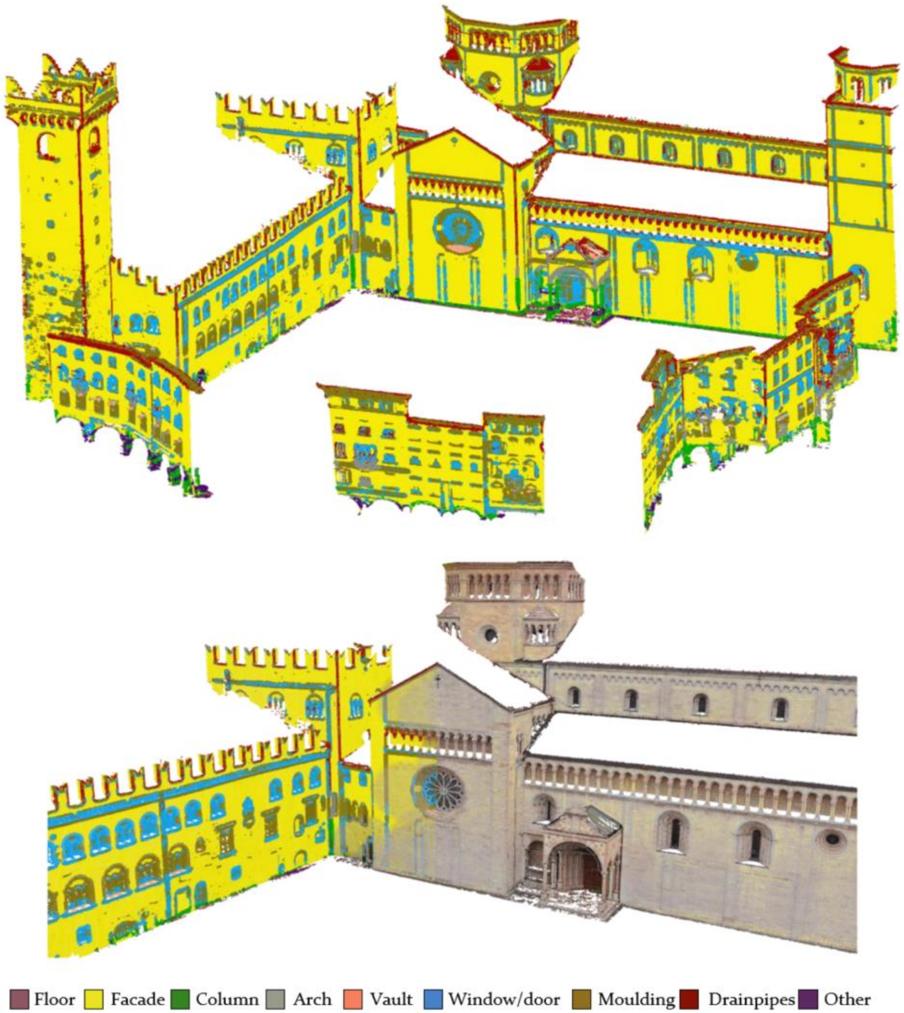


Image Source: Grilli, E., & Remondino, F. (2020). Machine Learning Generalisation across Different 3D Architectural Heritage. ISPRS International Journal of Geo-Information, 9(6), 379.

Examples [4]

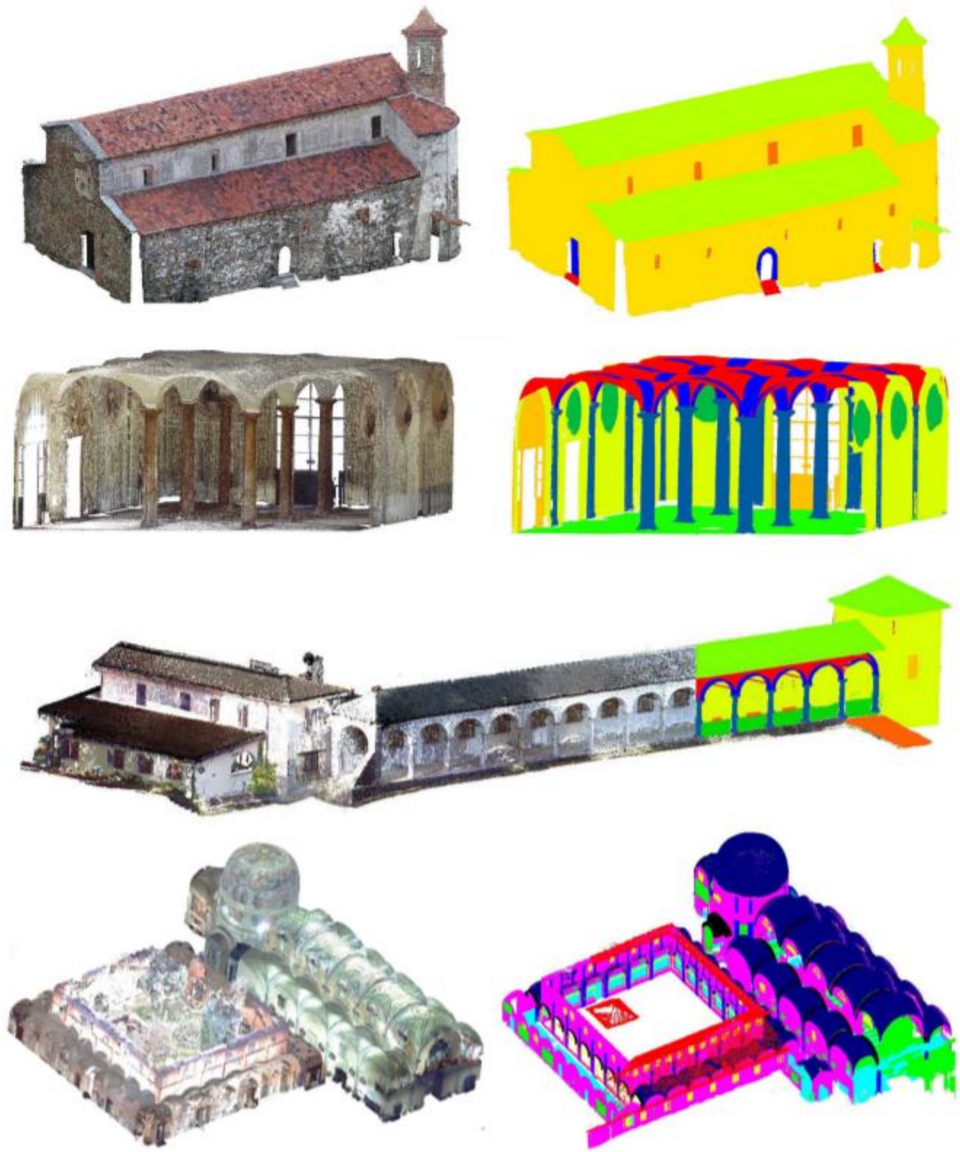
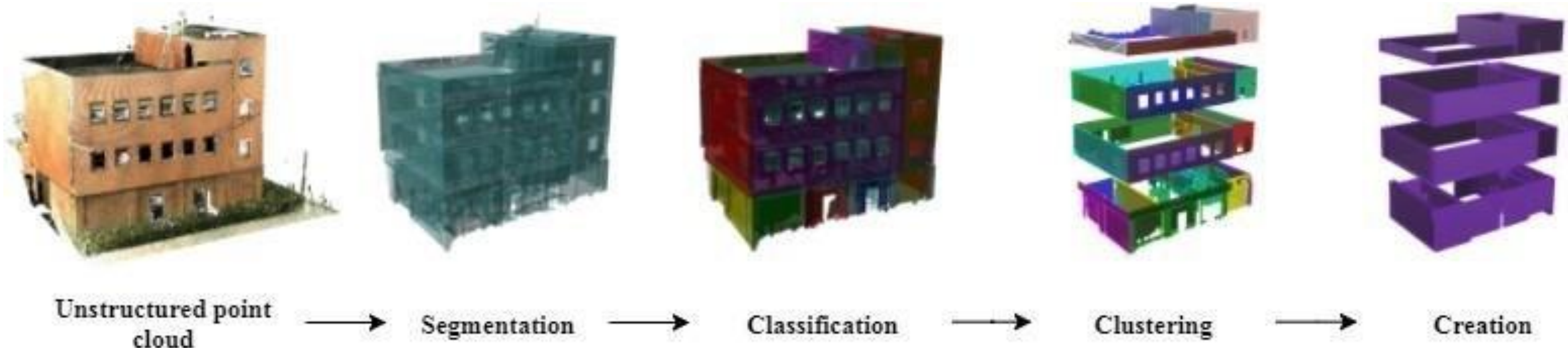


Image Source: Malinverni, E. S., Pierdicca, R., Paolanti, M., Martini, M., Morbidoni, C., Matrone, F., & Lingua, A. (2019). DEEP LEARNING FOR SEMANTIC SEGMENTATION OF 3D POINT CLOUD. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences.

Examples [5]

- Semantic classification for BIM



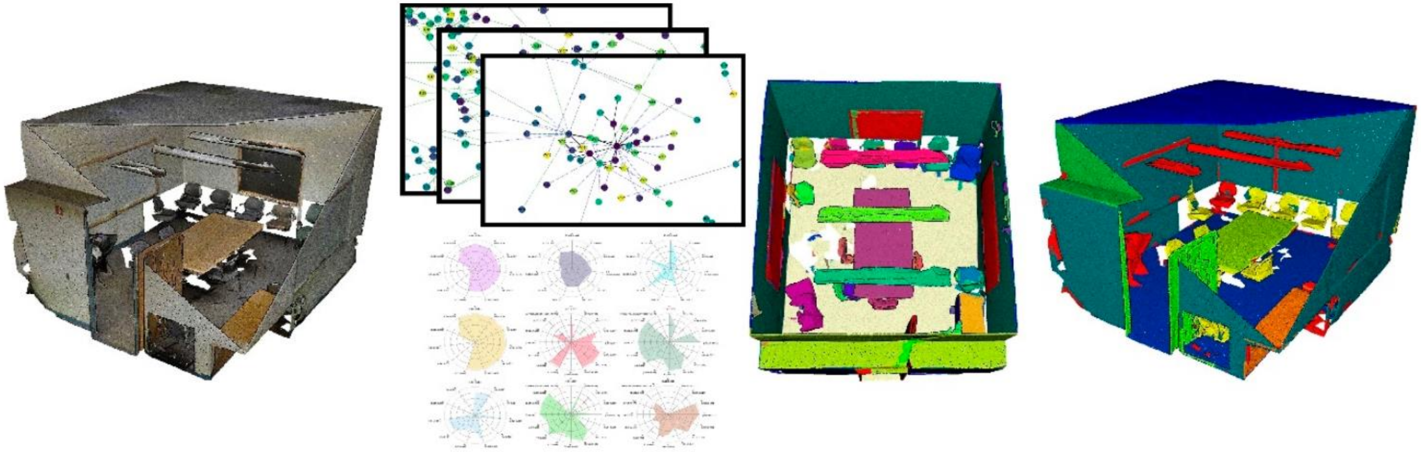


Image Source: (up) Bassier, M., Yousefzadeh, M., & Vergauwen, M. (2020). Comparison of 2D and 3D wall reconstruction algorithms from point cloud data for as-built BIM. Journal of Information Technology in Construction (ITcon), 25(11), 173-192.
(down) Poux, F., & Billen, R. (2019). Voxel-based 3D point cloud semantic segmentation: unsupervised geometric and relationship featuring vs deep learning methods. ISPRS International Journal of Geo-Information, 8(5), 213.

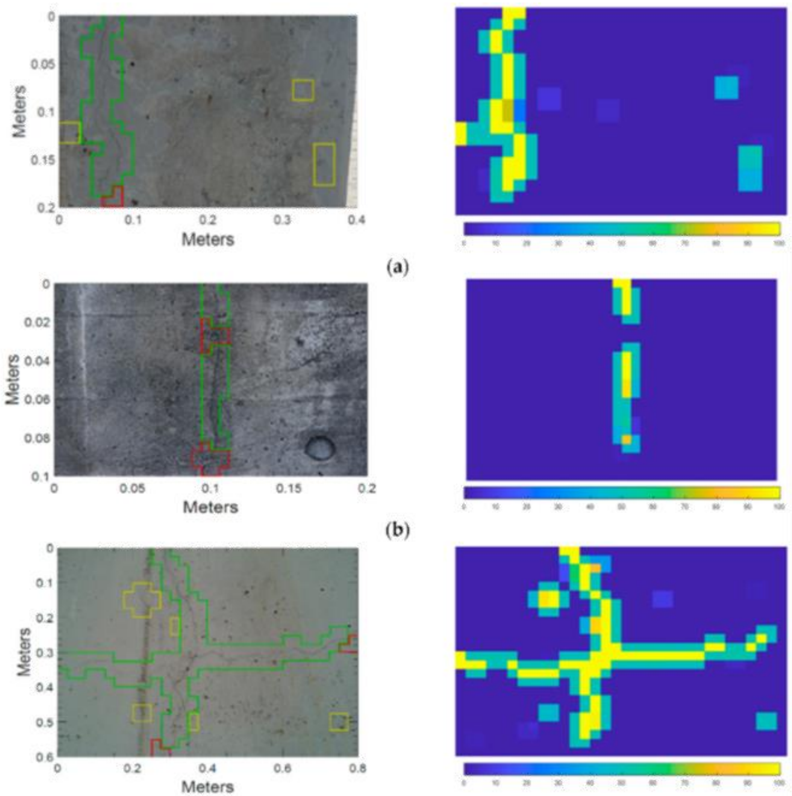
Examples [6]

- Semantic classification can be also extended to whole cities' 3D models

Image Source: http://www.gisresources.com/pix4d-announces-machine-learning-point-cloud-classification/

Examples [7]

- CNNs to detect cracks and perform automated inspection on orthoimages and photos



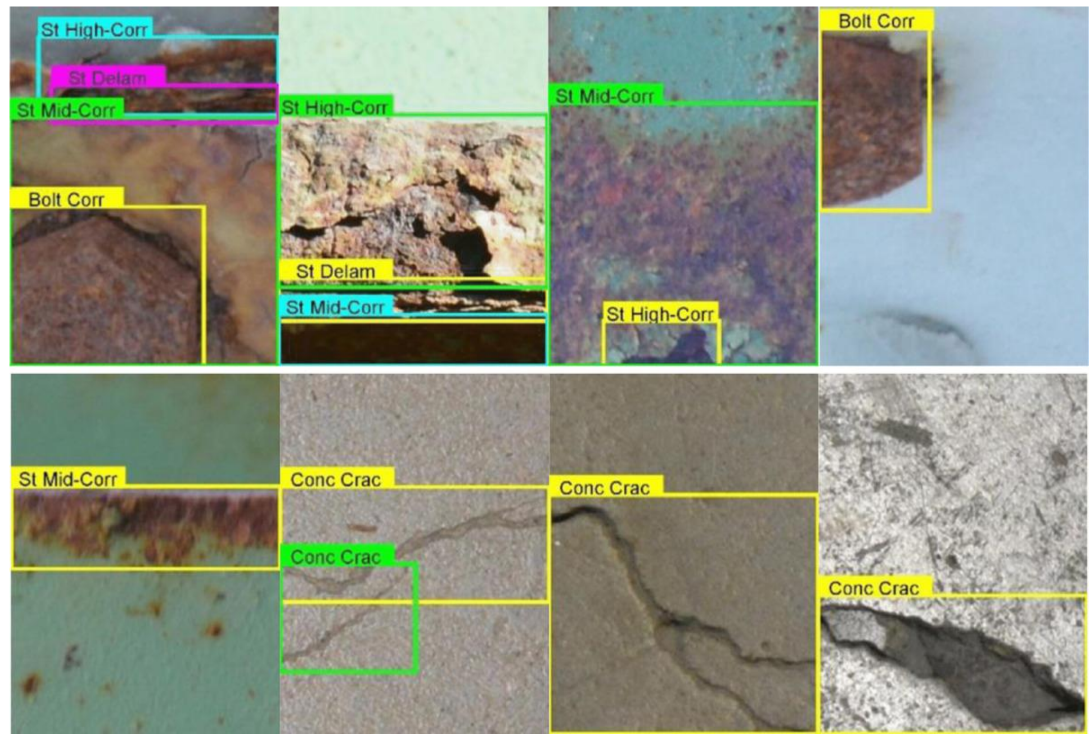
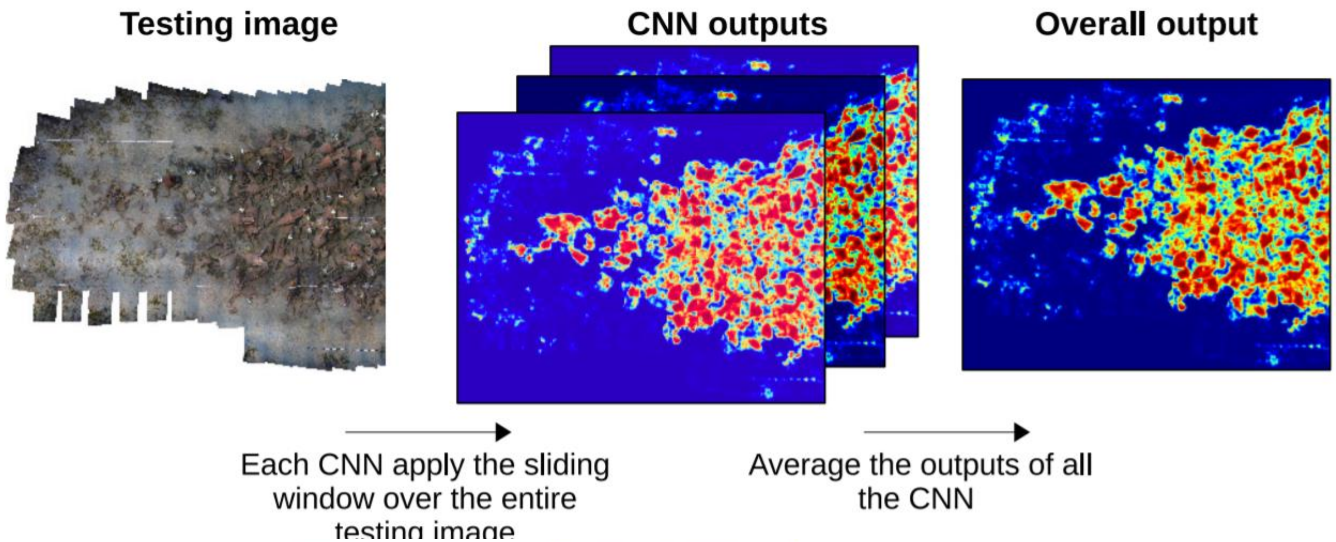


Image Source: Kim, B., & Cho, S. (2018). Automated vision-based detection of cracks on concrete surfaces using a deep learning technique. Sensors, 18(10), 3452 (left), Cha, Y. J., Choi, W., Suh, G., Mahmoudkhani, S., & Büyüköztürk, O. (2018). Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types. Computer-Aided Civil and Infrastructure Engineering, 33(9), 731-747. (right)

Examples [8]



- CNNs to detect and extract specific types of amphorae of an ancient shipwreck

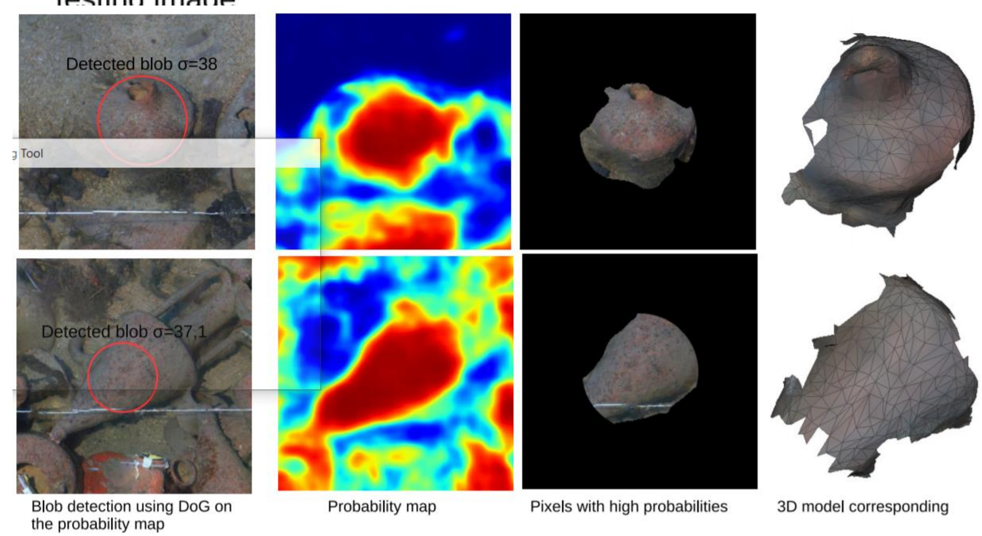


Image Source: Pasquet, J., Demesticha, S., Skarlatos, D., Merad, D., & Drap, P. (2017, October). Amphora detection based on a gradient weighted error in a convolution neuronal network. In IMEKO international conference on Metrology for Archaeology and Cultural Heritage, Lecce, Italy (pp. 23-25).

Notes!

- We can apply semantic classification the point cloud directly OR on the images
- **When it is applied on the 3D point clouds** may fail due to the sparsity of the points. Sometimes is more complicated.
- **When it is applied on the images**, the semantic information must be transferred to the 3D model through the photogrammetric process and the 3D information is not exploited
- However, **this way, the semantic information on the scene can be exploited in photogrammetry to increase the matching accuracy**
- There is not a straightforward way. It depends on the needs.

Critical reflections on the use of ML in CH

- The restricted access to data is an obstacle in allowing the trained networks to generalize
- Data must be of good quality, otherwise the accuracy of the results is deteriorated
- Machine learning frameworks are not black boxes, but mathematical and statistical functions
- They are not delivering the expected results when they are used as black boxes

Conclusions

- Photogrammetry and 3D computer vision serve as a valuable tool when studying the structural stability of historical buildings
- Image quality and camera positions are affecting the quality and the accuracy of the produced results
- The process is almost fully automated, enabling fast and cheap implementation
- Intermediate results must be evaluated by experts in order to monitor the process. The generation of a point cloud does not mean that it is correct.
- ML/DL will boost the offered information on the photogrammetric results, facilitating also more accurate results, when the semantic information is exploited

Thank you!

Questions?