

FROM 2D TO 3D SUPERVISED SEGMENTATION AND CLASSIFICATION FOR CULTURAL HERITAGE APPLICATIONS

E. Grilli^{a,b}, D. Dininno^{a,c}, G. Petrucci^{d,e,f}, F. Remondino^a

^a 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy

^b Alma Mater Studiorum, Bologna-(Italy)

^c Scienze dell'antichità e archeologia, University of Pisa (Italy)

^d Data and Knowledge Management (DKM) unit & Process and Data Intelligence (PDI) unit, Bruno Kessler Foundation (FBK), Trento, Italy

^e University of Trento, Italy

^f Google

<grilli><dininno><petrucci><remondino>@fbk.eu

Commission II/WG8

KEYWORDS: Classification, Segmentation, Cultural Heritage, Machine Learning, Random Forest

ABSTRACT:

The digital management of architectural heritage information is still a complex problem, as a heritage object requires an integrated representation of various types of information in order to develop appropriate restoration or conservation strategies. Currently, there is extensive research focused on automatic procedures of segmentation and classification of 3D point clouds or meshes, which can accelerate the study of a monument and integrate it with heterogeneous information and attributes, useful to characterize and describe the surveyed object. The aim of this study is to propose an optimal, repeatable and reliable procedure to manage various types of 3D surveying data and associate them with heterogeneous information and attributes to characterize and describe the surveyed object. In particular, this paper presents an approach for classifying 3D heritage models, starting from the segmentation of their textures based on supervised machine learning methods. Experimental results run on three different case studies demonstrate that the proposed approach is effective and with many further potentials.

1. INTRODUCTION

The generation of 3D data of heritage monuments, in form of point clouds or meshes, is transforming the approach that of researchers, archaeologists and curators use for the analysis of the findings. 3D models allow to perform morphological measurements, map degradation or annotate sites and structures directly on the virtual reconstruction of the studied objects.

Management of architectural heritage information is crucial for better understanding heritage data and for the development of appropriate conservation strategies. An efficient information management strategy should take into consideration three main concepts: segmentation, structuring the hierarchical relationships and semantic enrichment (Saygi et al., 2013). But the demand for automatic model analysis and understanding is ever increasing. Recent years have witnessed significant progress in automatic procedures for segmentation and classification of point clouds or meshes. Segmentation is the process of grouping point clouds or meshes into multiple homogeneous regions with similar properties, whereas classification is the step that labels these regions (Grilli et al., 2017). There are multiple studies related to the segmentation topic, mainly driven by specific needs provided by the field of application (building modelling, heritage documentation and preservation, robotics, etc.). Most of the segmentation algorithms are tailored to work with a 2.5D surface model assumption, coming for example from a LiDAR-based survey. Many algorithms require a fine-tuning of different parameters depending upon the nature of data and applications. The majority of these are supervised methods, where a training phase is mandatory and fundamental to guide the successive machine learning classification solution (Guo et al., 2014; Niemeyer et al., 2014; Xu et al., 2014; Weinmann et al., 2015; Hackel et al., 2016; Qi et al., 2016; Weinmann et al., 2017; Wang et al., 2018).

It is proved that complex real-world tasks require large training data sets for classifier training. Different benchmarks were proposed in the research community, with the "Large-Scale Point Cloud Classification Benchmark" (www.semantic3d.net) providing labelled terrestrial 3D point cloud data on which people can test and validate their algorithms. Until now there are no datasets for 3D heritage point cloud classification which would be sufficiently rich in both object representations and number of labelled points.

Considering the availability and reliability of segmentation methods applied to (2D) images and the efficacy of machine learning strategies, we present our work and methodology developed to assist heritage workers in the analysis of the finds, whose core consists in the 2D segmentation of the texture of 3D models.

1.1 Aim of the paper

The possibility to semantically annotate shape parts may have a relevant impact in several domains, like architecture and archaeology. Regarding the segmentation phase, the identification of different architectural components in point clouds and 3D meshes is of primary importance. Such operations can facilitate the study of heritage monuments and integrate heterogeneous information and attributes, useful to characterize and describe the surveyed object.

The presented research was motivated by the concrete need of archaeologists to identify and map constructive functions and materials of heritage structures. In order to address this need, we developed a method to (i) document and retrieve historical and architectural information, (ii) distinguish different building techniques (e.g. types of *Opus*, etc.) and (iii) recognize the presence of previous restoration works. Detection of such types of information in historic buildings with traditional methods, such as manual mapping or simple eye examination by an expert,

are considered time-consuming and laborious procedures (Corso et al., 2017). The aim of our research is to propose a more efficient technique for classification with reduced manual input. The structure of the paper is as follows: Section 2 reports the state of the art in heritage segmentation and classification, focussing on previous studies for restoration purpose and walls analysis. Section 3 gives an assessment of the developed 3D segmentation methodology. Section 4 presents the case studies and assessment approach whereas results are reported in Section 5. Finally, conclusions wrap up the paper, reporting challenges and a future vision to fulfil the gap in the field.

2. RELATED WORKS

Many experiments were carried out about the segmentation of heritage 3D data at different scales (Manfredini et al., 2008; Barsanti et al., 2017; Cipriani et al., 2017; Poux et al., 2017). Some works aim to define a procedure for the integration of archaeological 3D models with BIM (Saygi et al. 2013; De Luca et al. 2014). Sithole (2008) proposes an automatic segmentation method for detecting bricks in masonry walls working on the point clouds, assuming that mortar channels are reasonably deep and wide. Oses et al. (2014) classify masonry walls using machine learning classifiers, support vector machines and classification trees. Riveiro et al. (2016) propose an algorithm for the segmentation of masonry blocks in point cloud based on a 2.5D approach that creates images based on the intensity attribute of LiDAR systems.

Recently the combination of digital technologies such as laser scanning, photogrammetry and computer vision-based techniques and 3D geographic information systems (3D GIS) have made a considerable contribution for the conservation strategies of ancient buildings. This is proposed in the NUBES project developed by CNRS-MAP, where the 3D model is generated from 2D annotated images. In particular, the NUBES web platform (Stefani et al. (2014) allows the displaying and cross-reference of 2D mapping data on the 3D model in real time, by means of structured 2D layer-like annotations concerning stone degradation, dating and material.

In Campanaro et al. (2016) a similar example to our paper is given. They created a 3D management system for heritage structures by exploiting the combination of 3D visualization and GIS analysis. The 3D model of the building was originally split into architectural sub-elements (facades) in order to add colour information projecting orthoimages by means of planar mapping techniques (texture mapping). In our case, the idea of categorizing the 3D model using UV maps avoids the creation of many different orthoimages, a challenging step for complex scenarios.

3. METHODOLOGY

3.1 Overview

Starting from coloured 3D point clouds or textured surface models, our pipeline (Fig.1) consists of the following steps:

1. Creation and optimization of geometries, orthoimages (for 2.5D geometries) and UV maps (for 3D geometries) for the heritage structure under investigation. In our tests (Section 5) all products are generated from photogrammetric data.
2. Manual orthoimage or UV map segmentation and class identification to create ground truth and training data (section 3.2).
3. Supervised segmentation, starting from the training dataset, of all the orthoimages and UV map of the digital models (section 3.3);

4. Projection of the classification results from 2D to 3D object space by back-projection and collinearity model.

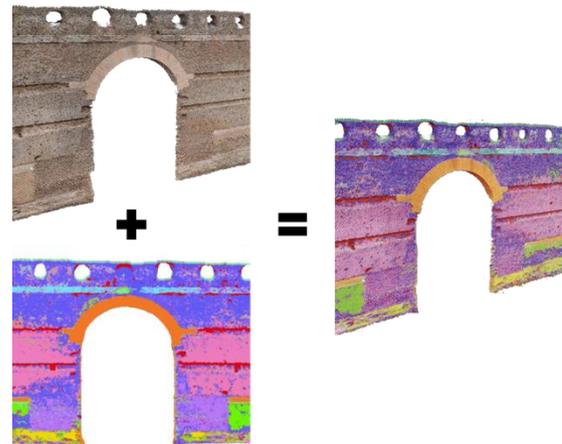


Figure 1: Schematic representation of the segmentation method.

3.2 UV map / texture generation

The innovative aspect of the presented method is that instead of working on many different 2D images or orthoimages generated from the 3D model, we decided to unwrap the textured 3D model and generate an UV map that can be classified with a supervised method. Firstly, to generate a good texture image to be classified, we followed these steps:

- Remeshing: it is useful to improve mesh quality and to facilitate the next steps;
- Unwrapping: UV maps are generated using Blender, adjusting and optimising seamlines and overlap (Fig.2a) to facilitate the subsequent analysis with machine learning strategies. This correction is made commanding the UV unwrapper to cut the mesh along edges chosen in accordance with the shape of the case study (Cipriani et al., 2017).
- Texture mapping: the created UV map is then textured (Fig. 2b) using the original textured polygonal model (as vertex colour or with external texture). This way the radiometric quality is not compromised despite the remeshing phase.

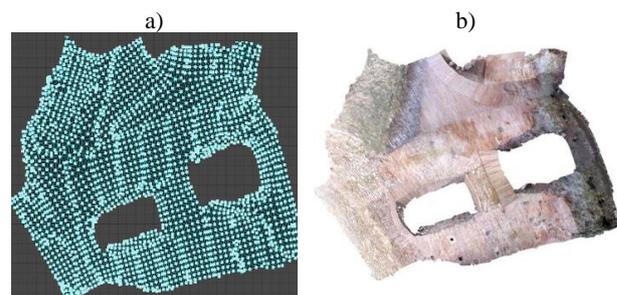


Figure 2: UV map after remeshing (a) and texturing (b) for the Cavea – Circus Maximus case study.

3.3 2D classification and segmentation

The 2D classification is performed using different machine learning models embedded in WeKa (Witten et al., 2016). Moreover we used the Fiji distribution of ImageJ, an image processing software that exploits WeKa as an engine for machine learning models (<http://imagej.net/Fiji>). These models are first trained by examples in a supervised way using a training set of manually annotated images.

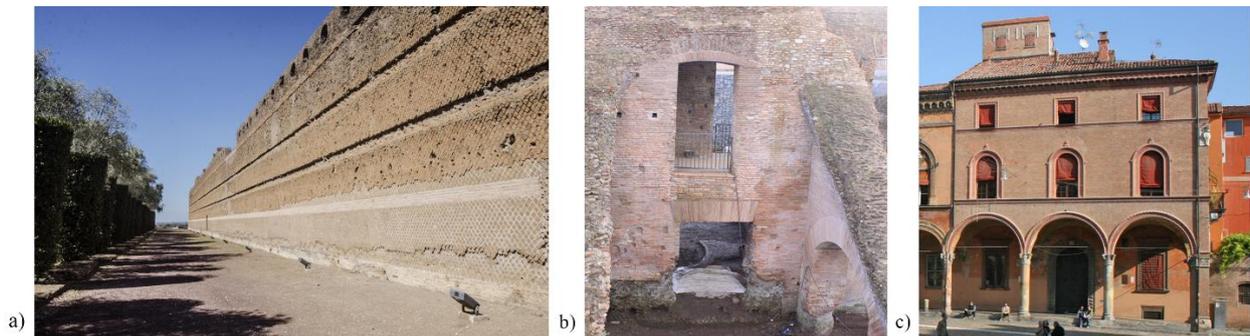


Figure 3: The case studies of the work to validate the semantic classification for analyses and restoration purposes: Pecile’s wall of Villa Adriana in Tivoli, Italy (a), *Cavea* walls of the *Circus Maximus* in Rome, Italy (b) and portico in Bologna, Italy (c).

In these images, each pixel has been manually annotated with its corresponding label. For each of these examples, the original image is submitted to the model that computes its actual response. The weights of the model are subsequently adjusted in order to minimize the difference between this response and the annotation that represents the expected response of the model. The performance of the model is assessed measuring the performance against another set of images, different from the ones used during the training phase, so that the capabilities of the model to generalize over unseen data can be effectively measured. The performances of different models trained on images at different scales are presented in section 4.2.

4. TEST OBJECTS AND EVALUATION METHOD

Object classification is a fundamental task in archaeology and heritage architecture although it is very important to have a clearly defined purpose and practical procedures when developing and applying classification methods. Traditional classifications started in the 19th century and are still being developed (Adams et al., 2007). The material classification is usually carried out by the operator, directly on pictures, as a precautionary phase to analyse the structural behaviour of a building and for historical analysis. Performing this operation manually is typically a costly and time-consuming process.

4.1 Case studies

The proposed methodology has been applied firstly to two different but coeval archaeological case studies, to verify the applicability of the methodology using a 2.5D and a 3D model: Villa Adriana in Tivoli, in particular focusing the attention on a portion of Pecile’s wall (60m L x 9m H) (Fig.3a) and a small portion of *cavea* walls of the *Circus Maximus* in Rome (5m L x 9m H x 2,5m D)(Fig. 3b). We classify the two digital models identifying the different categories of *Opus* (roman building techniques), distinguishing within the same class original and restored parts.

The last case study hereafter presented is part of a portico located in the city centre of Bologna, spanning ca 8m L x 13m H x 5m D (Fig. 3c).

4.2 Assessment methodology

In order to automatically assess the performance of the classification, we rely on the accuracy computed for each pixel by comparing the label predicted by the classifier with the same manually annotated. Then we compute the ratio between the number of correctly classified pixels by the total number of pixels as:

$$Accuracy = \frac{N^{\circ} \text{ of correct classifications}}{\text{total number of pixels}}$$

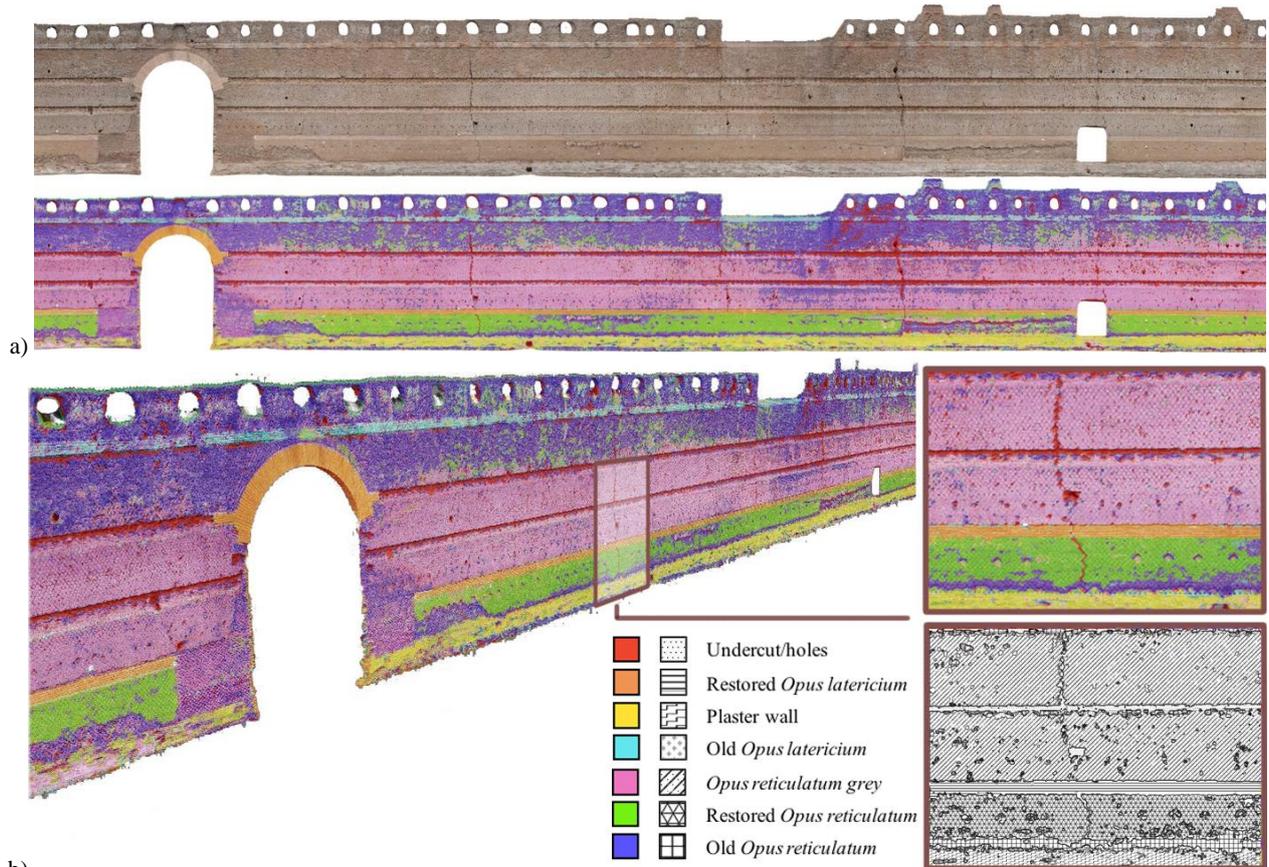
5. RESULTS

5.1 Pecile’s wall

Different training processes were run using different orthoimage scales (Fig. 4) in order to identify the best fitting solution for our case studies. With a 1:10 scale, we obtained results of over segmentation. Using a 1:50 scale, many details were lost, identifying only some macro-areas. The scale 1:20 (normally used for restoration purposes as it allows to distinguish bricks) turned out to be the optimal choice. It allows to capture the details but is still capable to not consider the cracks of the mortar between a brick and the other (Fig. 5). Given the manually selected training classes, we trained and evaluated different classifiers (Table 1). The first time the training process starts, the features of the input image will be extracted and converted to a set of vectors of float values (Weka input). This step can take some time depending on the size of the images, the number of features and the number of cores of the machine where the classification is running. The feature calculation is done in a completely multi-thread fashion. The features will be only calculated the first time we train after starting the plugin or after changing any of the feature options. Table 1 reports the accuracy results for all tested classifiers run on the orthoimage at scale 1.20. Moreover, we report the time elapsed for each algorithm, considering that creating the classes and the training data took around 10 minutes and the feature stack array took 14 minutes.

Classifier	Accuracy	Time
<i>j48</i>	0.44	22 s
<i>Random Tree</i>	0.46	15 s
<i>RepTREE</i>	0.47	33 s
<i>LogitBoost</i>	0.52	20 s
<i>Random Forest</i>	0.57	23 s
<i>Fast Random Forest (16)</i>	0.64	76 s
<i>Fast Random Forest (40)</i>	0.69	120 s

Table 1: Accuracy results and elapsed time for various classifier applied to an orthoimage at 1:20 scale.



b)
 Figure 6: The original and classified orthoimage of a longer part of the Pecile's wall long ca 60 m (a). Classification results mapped onto the 3D model of the wall (b). A closer view is also reported to better show the classification results with random colours or dedicated symbols.

5.2 Cavea of Circus Maximus

The unwrapping procedure (Section 3.2) of the complex 3D model of the *Cavea* allowed to classify the whole model without the need of generating multiple ortho-views. A more articulated training (Fig. 7) was created on the 1:20 UV map to classify the *Cavea*, choosing micro areas to identify the phases of intervention on the monuments (ancient and recent restorations, integrations and changes of the building arrangements). The manual training took ca 30 min.

The classification results (achieved using Fast Random Forest method) are quite satisfactory (Fig. 8). The algorithm could easily recognize within the same categories of *opus* three different types of restoration and was also able to identify the *opus reticulatum* class, even though it occupied only a small and dark portion of the object. This type of result underlines the quality of details that can be achieved starting from a detailed manual training.

The visualization of the classification results on the 3D geometry allow heritage end-users to even see restorations located in undercuts. The segmentation made with such a level of detail is useful for mapping the deterioration and to calculate the volumes for planning future restorations.



Figure 7: Manually identified training areas (11 classes) on the unwrapped texture of the Circus's cavea.

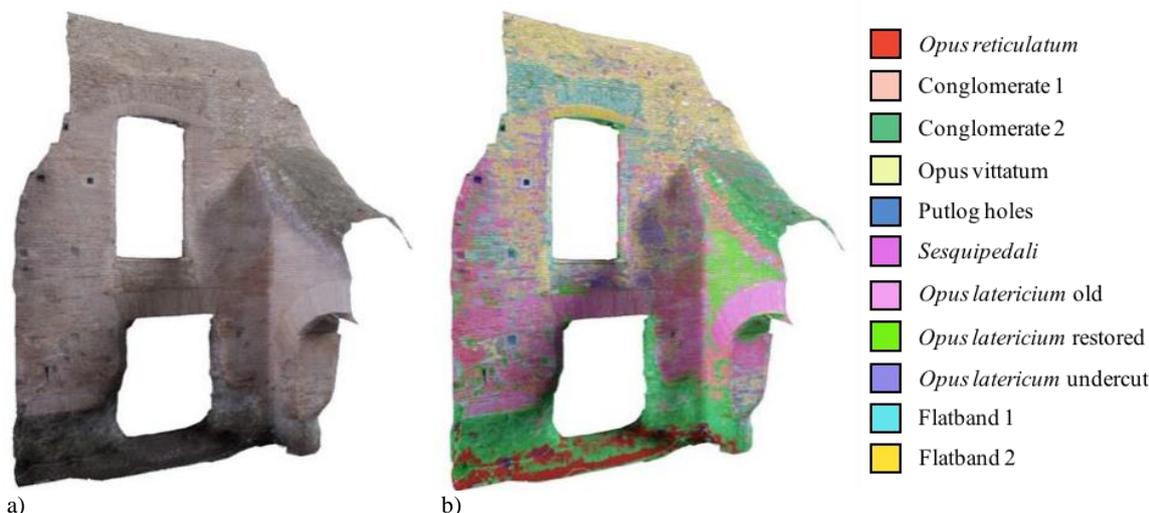


Figure 8: Original and segmented 3D model of a portion of Roman Cavea in the *Circus Maximus* in Rome (Italy).

5.3 Classification of portico structures

To demonstrate the replicability of the proposed method to a different type of 3D model, a third case study featuring a historical portico dataset is used. Such structures combine variegated geometric shapes, different materials and many architectural details like mouldings and ornaments. According to the different classification requirements, the aim of the task could be:

- Identification of construction techniques;
- Identification of different materials (bricks vs stones vs marble);
- Identification of degradation categories (cracks vs humidity vs swelling).

Figure 9 show the unwrapped texture of the photogrammetric 3D model with the manually identified training patches and classes (11). We decided to split some categories (walls and columns) into two different classes to prevent error moved by shadows and different plaster chromatism. The classification results (Fig. 10), based on Fast Random Forest model / classifier, show an over segmentation under the porticoes, where the plaster of the wall is not homogeneous and presents different types of degradation. In this case a solution might be to create many different classes according to the number of degradation categories or apply an algorithm to homogenise areas with small spots.



- Door
- Wall 1
- Wall 2
- Window
- Pavement
- Vaults
- Ornamental elements
- Column capitals and bases
- Column Shaft
- Column Shaft (shade)
- Doorframe

Figure 9: Manual training classes to segment the building for semantic purposes.

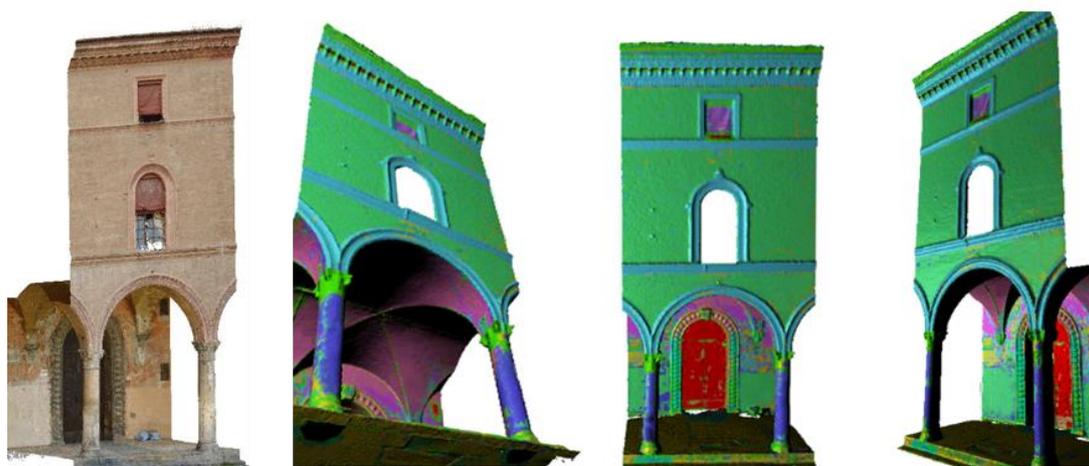


Figure 10: 3D model and classification results of an historical building in the city centre of Bologna.

5. CONCLUSIONS

With the proposed methods, archaeologists or curator specialists are able to automatically annotate 2D textures of heritage objects and visualize them onto 3D geometries for a better understanding. The difficulty of applying image segmentation to cultural heritage case studies derives firstly from the existence of a large amount of building techniques and ornamental elements. A monument can be subjected to different types of degradation, according to the different conditions, hence increasing the efficiency of the classification tasks. A machine learning-based approach becomes beneficial for speeding up classification tasks on large and complex scenarios, provided that training datasets are as much differentiated as possible.

In summary, critical issues and possible solutions are:

- dark holes in architecture structures (e.g. Putlog holes) can be confused with shadows introducing errors in the classification: it becomes fundamental the choice of the right classes during the training phase;
- over-segmentation provides too many classes not useful in case of semantic analysis: it is necessary to apply some algorithm to homogenise regions;
- long and time-consuming training phase, in particular in case of many classes or non-homogeneous surface (see Fig.7): for a better results a detail training it's necessary.

On the other hand, the advantages of the proposed method are:

- shorter time to classify objects wrt manual methods (see Table 1);
- over-segmentation useful for restoration purposes to detect small cracks or deteriorated parts;
- the training set might be used (replicated) for buildings of the same historical period or with similar construction material;
- using unwrap texture allows the visualization of classification results onto 3D models from different points of view;
- the pipeline can be extended to different kinds of heritage buildings, monuments or 3D models in general.

As a future work, we plan to exploit more complex machine learning algorithms, in particular Deep Neural Networks to learn more expressive representations of the image. In particular, we will tackle the objective of increasing the homogeneity of the segmentation in order to minimize, and ideally avoid, any post-processing phase. In order to train such models, a larger amount of training examples is needed. Hence, more effort will be put in the activity of extending the training set with more manually annotated examples.

ACKNOWLEDGEMENTS

The authors acknowledge Accademia Adrianea and University of Bologna (Dept. Architecture) for providing the datasets of Pecile's wall in Villa Adriana and Bologna's porticoes. We want to thank also Soprintendenza Capitolina for the permission to survey the *Circus Maximus* in Rome.

REFERENCES

Adams, W. Y., Adams, E. W., 2007: *Archaeological typology and practical reality: a dialectical approach to artefact classification and sorting*. Cambridge University Press.

Apollonio, F. I., Basilissi, V., Callieri, M., Dellepiane, M., Gaiani, M., Ponchio, F., Rubino A., Scopigno, R., 2017: A 3D-centered information system for the documentation of a complex

restoration intervention. *Journal of Cultural Heritage*, Vol. 29, pp 89.

Barsanti, S. G., Guidi, G., & De Luca, L., 2017: Segmentation of 3D models for cultural heritage structural analysis-some critical issues. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, Vol. IV(2-W2), pp 115-122.

Breiman, L., 2001: Random forests. *Machine learning*, Vol. 45(1), 5-32.

Campanaro, D. M., Landeschi, G., Dell'Unto, N., & Touati, A. M. L. 2016: 3D GIS for cultural heritage restoration: A 'white box' workflow. *Journal of Cultural Heritage*, Vol. 18, pp.321-332.

Cipriani, L., Fantini, F., Bertacchi, S., 2014: 3D models mapping optimization through an integrated parameterization approach: cases studies from Ravenna. *ISPRS International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 40(5), p.173.

Cipriani, L., Fantini, F., 2017: Digitalization culture VS archaeological visualization: integration of pipelines and open issues. *ISPRS International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 42(2-W3), pp 195-202.

Corso, J., Roca, J., Buill, F., 2017: Geometric analysis on stone façades with terrestrial laser scanner technology. *Geosciences*, 7(4), 103.

De Luca, L., & Buglio, D. L., 2014: Geometry vs Semantics: open issues on 3D reconstruction of architectural elements. In *3D Research Challenges in Cultural Heritage*, pp. 36-49. Springer, Berlin, Heidelberg.

Grilli, E., Menna, F., & Remondino, F., 2017: A review of point clouds segmentation and classification algorithms. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 42(2-W3), pp 339-344.

Guo, B., Huang, X., Zhang, F. and Sohn, G., 2014: Classification of airborne laser scanning data using JointBoost. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 92, pp. 124-136.

Hackel, T., Wegner, J.D., Schindler, K., 2016: Fast semantic segmentation of 3d point clouds with strongly varying density. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. III(3), pp. 177-184.

Manferdini, A. M., Remondino, F., Baldissini, S., Gaiani, M., & Benedetti, B., 2008: 3D modeling and semantic classification of archaeological finds for management and visualization in 3D archaeological databases. In *Int. Conference on Virtual Systems and MultiMedia (VSMM)*, pp. 221-228.

Manuel, A., De Luca, L., Véron, P., 2014: A hybrid approach for the semantic annotation of spatially oriented images. *International Journal of Heritage in the Digital Era*, Vol. 3(2), pp. 305-320.

Niemeyer, J., Rottensteiner, F. and Soergel, U., 2014: Contextual classification of lidar data and building object detection in urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 87, pp. 152-165.

Oses, N., Dornaika, F., & Moujahid, A., 2014: Image-based delineation and classification of built heritage masonry. *Remote Sensing*, 6(3), pp 1863-1889.

Poux, F., Neuville, R., Hallot, P., & Billen, R., 2017: Point cloud classification of tesserae from terrestrial laser data combined

with dense image matching for archaeological information extraction. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 4, 203-211.

Qi, C. R., Su, H., Mo, K., & Guibas, L. J., 2017: Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proc. Computer Vision and Pattern Recognition (CVPR)*, IEEE, 1(2), 4.

Riveiro, B., Lourenço, P. B., Oliveira, D. V., González-Jorge, H., & Arias, P.: 2016. Automatic Morphologic Analysis of Quasi-Periodic Masonry Walls from LiDAR. *Computer-Aided Civil and Infrastructure Engineering*, 31(4), pp. 305-319.

Saygi, G.; Remondino, F. 2013: Management of Architectural Heritage Information in BIM and GIS: State-of-the-art and Future Perspectives. *International Journal of Heritage in the Digital Era*, Vol. 2(4), pp. 695-713.

Scopigno, R., Callieri, M., Cignoni, P., Corsini, M., Dellepiane, M., Ponchio, F., Ranzuglia, G. 2011: 3D models for cultural heritage: Beyond plain visualization. *Computer*, 44(7), pp.48-55.

Sithole, G., 2008: Detection of bricks in a masonry wall. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 37 (5), pp. 1-6.

Stefani C., Busayarat C., Lombardo J., De Luca L., 2013: A web platform for the consultation of spatialized and semantically enriched iconographic sources on cultural heritage buildings. *International Journal on Computing and Cultural Heritage (JOCCH)*, Vol.6(3).

Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., & Solomon, J. M., 2018: Dynamic Graph CNN for Learning on Point Clouds. arXiv preprint arXiv:1801.07829.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J., 2016: *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Weinmann, M., Schmidt A., Mallet C., Hinz S., Rottensteiner F., Jutzi B., 2015: Contextual classification of point cloud data by exploiting individual 3D neighbourhoods. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. II(3-W4).

Weinmann, M., & Weinmann, M., 2017: Geospatial Computer Vision Based on Multi-Modal Data—How Valuable Is Shape Information for the Extraction of Semantic Information? *Remote Sensing*, 10(1), 2.

Xu, S., Vosselman, G. and Oude Elberink, S., 2014: Multiple-entity based classification of airborne laser scanning data in urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 88, pp. 1-15.