

Alma Mater Studiorum – Università di Bologna

**DOTTORATO DI RICERCA IN
ARCHITETTURA**

Ciclo XXXII

Settore Concorsuale: 08/E1 – DISEGNO

Settore Scientifico Disciplinare: ICAR 17 Disegno

**“Automatic classification of architectural and
archaeological 3D Data”**

Presentata da: Grilli Eleonora

Coordinatore Dottorato

Prof.sa Annalisa Trentin

Supervisor

Dr. Prof. Luca Cipriani

Co-Supervisor

Dr. Fabio Remondino

Esame finale anno 2020

CONTENT

PREFAZIONE.....	1
PREAMBLE.....	3
PART 1 – INTRODUCTION AND STATE OF THE ART	
1. 3D technologies for Cultural Heritage.....	9
2. An overview of segmentation and classification methods.....	17
2.1 Traditional segmentation approaches.....	18
2.1.1 Edge-based segmentation.....	18
2.1.2 Region-based segmentation.....	19
2.1.3 Segmentation by model fitting.....	20
2.2 Machine learning classification approaches.....	22
2.2.1 Support Vector Machine.....	23
2.2.2 Decision trees and Random Forest.....	24
2.2.3 K-means clustering	25
2.2.4 Artificial neural networks.....	26
2.3 Classification strategies applied to Cultural heritage.....	27
3. Aim of the research.....	37
PART 2 – PROPOSED METHODS	
4. Textured-based classification approach.....	43
4.1 From 3D to 2D.....	44
4.1.1 Orthophotos.....	44
4.1.2 UV maps.....	44
4.2 Colour spaces.....	46
4.3 Supervised learning classification.....	46

Content

4.4 Unsupervised learning segmentation.....	48
4.5 Surfaces computation.....	48
4.5.1 Quantity survey validation.....	48
5. Geometry-based classification approach.....	53
5.1 Manual annotation.....	54
5.2 Feature extraction.....	54
5.2.1 The covariance features.....	55
5.3 Model training.....	57
5.3.1 Random Forest classifier.....	57
5.3.2 One versus One classifier.....	57
5.3.3 1D and 2D Convolutional Neural Network.....	58
5.3.4 Bi-LSTM classifier.....	58
5.4 Validation.....	58
5.5 Feature selection.....	60
5.6 Computational Geometry Algorithms Library (CGAL)	62
PART 3 – CASE STUDIES	
6. Textured-based classification results.....	69
6.1 Pecile wall in Villa Adriana, Tivoli	70
6.2 Sarcophagus of the Spouses, Villa Giulia museum, Rome.....	74
6.3 Bartoccini's tomb, Tarquinia (1).....	77
6.4 Porticoes in Bologna (1).....	81
7. Geometry-based classification results.....	85
7.1 Bartoccini's tomb, Tarquinia (2).....	86
7.2 Basilica, Paestum.....	88

Content

7.2.1	Assessment of the feature selection workflow.....	89
7.3	Ad hoc features reliability.....	95
7.4	Temple of Neptune, Paestum.....	96
7.5	Porticoes in Bologna (2).....	102
7.5.1	Classification using CGAL	102
7.5.2	Classification using ad hoc features.....	103
CONCLUSION AND FUTURE WORKS.....		111
AKNOWLEDGMENTS.....		114

PREFAZIONE

Il settore dei beni culturali, negli ultimi decenni, è stato uno dei campi di sperimentazione più interessanti delle nuove tecnologie e metodologie di rilievo digitale tridimensionale. L'evoluzione digitale nel settore dei beni culturali ha profondamente trasformato le modalità di acquisizione, elaborazione e gestione dei dati. L'applicazione delle nuove metodologie di rilievo e documentazione tridimensionale, attraverso l'utilizzo di sensori attivi e passivi, è ormai una pratica comune. Queste tecnologie di rilievo garantiscono, in tempi ridotti, la produzione di nuvole di punti e modelli 3D dotati di livelli di precisione molto elevati. La diffusione di questo tipo di dati è in continua crescita, divenendo così elementi facilmente condivisibili sul web e persino accessibili su dispositivi mobili quali smartphone e tablet. Al fine di sfruttare le reali potenzialità di questa significativa quantità di dati e poter estrarre informazioni semantiche da nuvole di punti o modelli poligonali, emerge oggi la necessità di sviluppare metodi affidabili di classificazione, che permettano di conferire una connotazione agli oggetti rappresentati in 3D. In generale, con il termine classificazione o segmentazione semantica si fa riferimento al processo di raggruppamento di dati simili in sottoinsiemi detti segmenti. Questi ultimi hanno in comune una o più caratteristiche (geometriche, radiometriche, ecc.), attraverso cui è possibile distinguere ed identificare le diverse parti che compongono un'immagine, una nuvola di punti o un modello poligonale. Date le grandi dimensioni dei dati raccolti, l'utilizzo di algoritmi automatici è sicuramente preferibile alle lunghe e tediose procedure di annotazione manuale.

Considerando inoltre la complessità e varietà delle nuvole di punti, che si differenziano, a seconda dell'oggetto rilevato, per densità, distribuzione dei punti, valori RGB etc., la ricerca è ancora in costante evoluzione. Sebbene con l'avvento dell'Intelligenza Artificiale siano emersi progressi significativi nelle procedure automatiche di classificazione applicate al campo geo-spaziale, ad oggi nulla è stato implementato in ambito architettonico.

Prefazione

Nel campo dei beni culturali, l'identificazione all'interno di nuvole di punti o modelli poligonali di vari elementi architettonici, piuttosto che la distinzione di materiali o stati di conservazione, può diventare uno strumento di studio prezioso. Tuttavia, se si considera la grande varietà e complessità dello studio in ambito archeologico/architettonico, l'applicazione di procedure di classificazione automatica diventa un compito particolarmente impegnativo.

Partendo da tali premesse, il percorso di dottorato è stato focalizzato allo sviluppo e alla convalida di procedure affidabili e automatizzate per la classificazione di dati architettonici e archeologici 3D (nuvole di punti o modelli poligonali provenienti da elaborazioni fotogrammetriche o rilievi laser scanner).

Attraverso i due diversi approcci sviluppati è stato possibile:

- Distinguere all'interno di un modello architettonico diverse tecniche costruttive (Sezione 6.1);
- Evidenziare i restauri esistenti (Sezione 6.2);
- Quantificare diversi stati di conservazione e materiali (Sezione 7.1);
- Identificare e distinguere elementi architettonici strutturali e decorativi (Sezione 7.2 - Sezione 7.5).

La tesi è divisa in tre diverse parti:

- Parte 1: in questa sezione viene fornita un'introduzione all'argomento, insieme ad un'ampia panoramica della letteratura sulle tecniche di classificazione 2D e 3D, applicate ai diversi ambiti (medicina, robotica, urbanistica etc.).
- Parte 2: la seconda parte della tesi è interamente dedicata alla descrizione dei due diversi approcci di classificazione sviluppati durante il dottorato. In particolare, l'approccio presentato al Capitolo 4 attiene ad un'attività di lavoro sulle texture dei modelli, mentre quello descritto al Capitolo 5 opera con la geometria della nuvola di punti.
- Parte 3: per verificare l'efficacia degli approcci sviluppati, nella terza parte della tesi viene trattata una vasta gamma di casi studio.

Tra i contributi che la ricerca ha apportato è importante annoverare:

- l'utilizzo di metodi di Machine Learning per la classificazione automatica di dati 3D riguardanti beni culturali;
- lo sviluppo di due diversi approcci, uno basato sull'uso della texture e l'altro sulla geometria, in modo da potersi muovere all'interno dell'ampio spettro di scenari appartenenti al patrimonio culturale;
- l'applicazione dei metodi proposti ad un ampio set di dati, al fine di dimostrarne la replicabilità;
- il raggiungimento di risultati di classificazione utili per l'interpretazione, il monitoraggio e il restauro dell'architettura, le applicazioni HBIM, ecc.

PREAMBLE

In the last years, various research activities were motivated by the need for systems and methodologies for the collection, management and processing of archaeological and architectural 3D data. Thanks to the advances in data acquisition technologies, that guarantee high levels of precision, extraordinary speed and versatility, the amount of available 3D dataset has exponentially increased. Point clouds and 3D models have become commonly shared data through the web, and accessible on mobile platforms. However, to fully explore the potential that such precise representation offers, especially in the heritage field of research, sometimes it becomes necessary annotating the models, providing them with meaningful attributes that give a connotation to the objects represented in 3D.

Given the large size of the collected point clouds, automatic algorithms are preferable to reduce the necessity of expensive and slow human processing in annotating procedures. In this context, automatic segmentation and classification approaches become paramount. Incidentally, segmentation is the process of grouping the data into multiple homogeneous regions with similar properties, whereas classification is the step that labels these regions.

Due to the complexity and variety of point clouds, caused by irregular sampling, varying density, different types of objects, etc., point cloud segmentation and classification are very active research topics. Although the revival of the Artificial Intelligence led to significant progress in automatically classifying 3D data in the geospatial field, its application to the architectural environment is still mainly unexplored.

In the cultural heritage field, the identification within point clouds or meshes of various architectural elements rather than the distinction of materials or states of preservation can become valuable tools for studying the objects with different purposes. However, if we consider the wide variety and complexity that heritage case studies feature, the application of automatic classification procedures becomes a really challenging task.

Preamble

Within this research landscape, the main goal of the PhD was to develop, test and validate reliable and automatic procedures for the classification of architectural and archaeological 3D data (point clouds or polygonal mesh models coming from photogrammetric processing or laser scanning surveying).

To achieve this goal, two different approaches have been developed based on texture or geometric information. They allow to:

- characterize different constructing techniques (Section 6.1);
- detect existing restoration evidence (Section 6.2);
- quantify different states of conservation and materials (Section 7.1);
- identify and distinguish structural and decorative architectural elements (Section 7.2 – Section 7.5).

This thesis is divided into three different parts:

- Part 1: in the first part an introduction to the topic is given, together with an extensive overview of the literature about 2D and 3D segmentation and classification techniques applied to different fields (Chapter 1-2-3).
- Part 2: the second part of the thesis is fully dedicated to the introduction of the two different classification methodologies developed during the PhD. In particular, the methodology presented in Chapter 4 works on the 2D textures of the models, while the one described in Chapter 5 operates with the geometry of the point cloud (Chapter 4-5).
- Part 3: to verify the effectiveness of the developed approaches, a wide range of case studies is treated in the third part of the thesis (Chapter 6-7).

The main contributions of the work are:

- the use of Machine Learning methods for the automated classification of 3D heritage data;
- the development of texture- and geometry-based approaches in order to consider the large spectrum of heritage scenarios and needs;
- the evolution of the proposed solution on a large set of data in order to prove their replicability;
- the achievement of classification results useful for interpreting architectures, monitoring and restoration purposes, HBIM applications, etc.

PART 1

Introduction and State of the art



CHAPTER 1

3D technologies for Cultural Heritage

According to the UNESCO guidelines Cultural heritage assets can be divided into tangible and intangible heritage. '*Intangible Cultural Heritage*' indicates '*the practices, representations, expressions, knowledge, skills – as well as the instruments, objects, artefacts and cultural spaces associated therewith – that communities, groups and, in some cases, individuals recognize as part of their Cultural Heritage*'. '*Tangible Cultural Heritage*' refers to *physical artefacts produced, maintained and transmitted intergenerationally in a society* (UNESCO, 2003). It includes buildings and historical places, monuments, artifacts, etc., which are considered worthy of preservation for the future. Tangible and intangible heritage require different approaches for preservation and safeguarding. In this research the focus will be only on tangible heritage.

The documentation and preservation of tangible heritage becomes fundamental for protecting and preserving it from armed conflicts, climate change effects or other natural catastrophes, and human-caused disasters. The presence of this risk is further enlarged by the fact that all the artifacts are inevitably in a constant state of chemical transformation. Even what is considered to be preserved is actually changing (Lowenthal, 2015).

The advent in the last decades of 3D optical instruments for the 3D digitization of objects and sites has undoubtedly changed the concept of heritage conservation and preservation. Indeed, the Cultural Heritage field is taking great advantage of reality-based surveying techniques (e.g. photogrammetry, laser scanning) (Gruen, 2008; Remondino, 2011). Currently, digital photogrammetry

Part 1. Introduction and State of the art

and laser scanning have become standard methods for data acquisition and digital recording for the 3D documentation of heritage assets. These technologies for 3D documentation allow the generation of realistic 3D results in terms of geometric and radiometric accuracy, overcoming the so-called direct surveys, which involve measuring in direct contact of objects or excavation areas.

The American Society for Photogrammetry and Remote Sensing (ASPRS) has defined Photogrammetry as *the art, science, and technology of obtaining reliable information about physical objects and the environment through processes of recording, measuring and interpreting photographic images and patterns of recorded radiant electromagnetic energy and other phenomena*. In contrast, laser scanning is the process of capturing digital information about the shape of an object with equipment that uses lasers to measure the distance between the scanner and the object. Once data are acquired (images, scans, single points, etc.), post-processing operations allow derivation of dense point clouds, polygonal models, orthoimages, sections, maps and drawing or further products. Towards providing precise representations of the objects at a given time to be passed down to future generations, these kinds of data can be used as a base for any further studies (Barsanti et al., 2014).

Among the different application:

- archaeological documentation (Guidi et al., 2014; Cipriani et al., 2017) (Figure 1.1);



Figure 1.1. Digital model of the Maritime theatre in Villa Adriana, Tivoli (Cipriani et al., 2017).

- digital preservation and conservation (Gomes et al., 2014) (Figure 1.2);



Figure 1.2. 3D model of an Indonesian board sculpture (Gomes et al., 2014).

- monitoring and deformation analysis (Abate, 2014; Chiabrandi et al., 2017) (Figure 1.3);

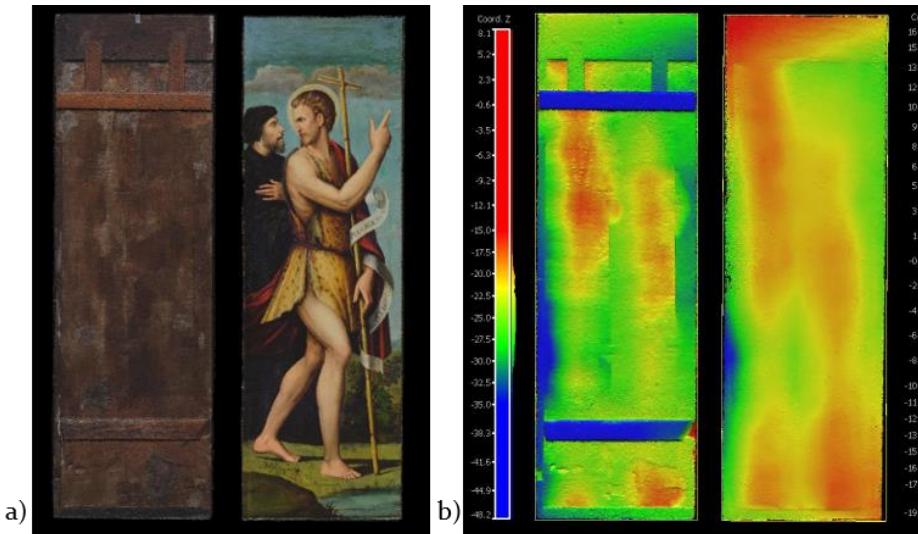


Figure 1.3. Analysis of the Saint John the Baptist painting: coloured dense point clouds (**a**) and depth map (**b**) (Abate, 2014).

- virtual reality/computer graphics applications (Fernández-Palacios et al., 2017; Bekele et al., 2018) (Figure 1.4);



Figure 1.4. VR application deployed for the immersive access of the Paestum Virtual Tour (Fernández-Palacios et al., 2017).

- computer-aided restoration (Apollonio et al., 2018; Roussel et al., 2019) (Figure 1.5);



Figure 1.5. 3D annotations onto 3D models: the Neptune's fountain in Bologna (Apollonio et al., 2018) (a); the Autumn statue in Marseille (Roussel et al., 2019)(b).

Part 1. Introduction and State of the art

- geographic web systems (Remondino et al., 2011; Baik et al., 2015) (Figure 1.6);



Figure 1.6. Geo-browser which allow the visualization of information related to the Etruscan civilization (Remondino et al., 2011).

- multimedia museum exhibitions (Callieri et al., 2015; Guidi et al., 2015) (Figure 1.7);

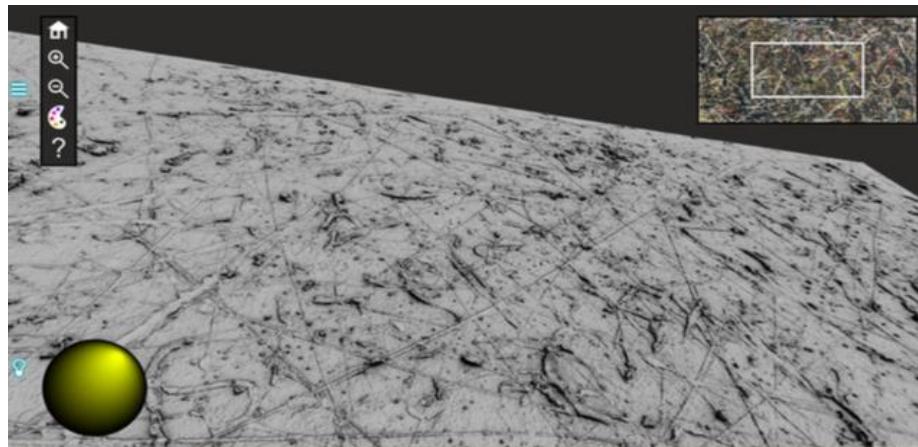


Figure 1.7. Interactive exhibition of the painting "Alchimia" by Jackson Pollock (Callieri et al., 2015).

- underwater documentation (Menna et al., 2018) (Figure 1.8);

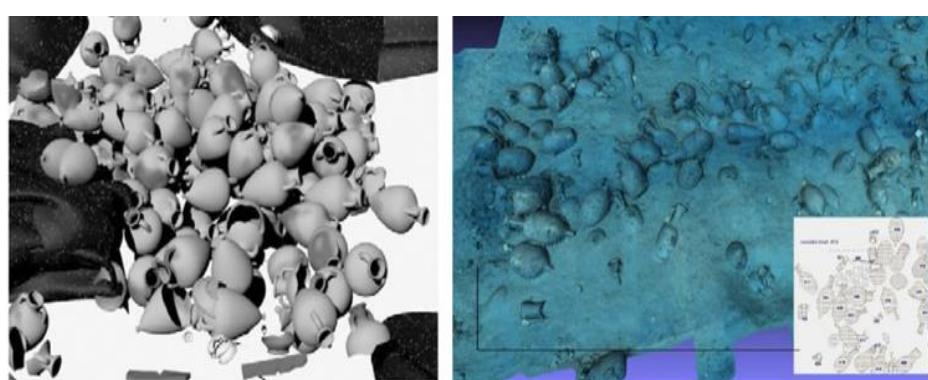


Figure 1.8. Examples of digital products obtained through photogrammetric acquisition carried out by scuba divers (Menna et al., 2018).

Chapter 1. 3D technologies for Cultural Heritage

- 3D repositories and catalogues (www.potree.org; www.3d-virtualmuseum.it) (Figure 1.9);

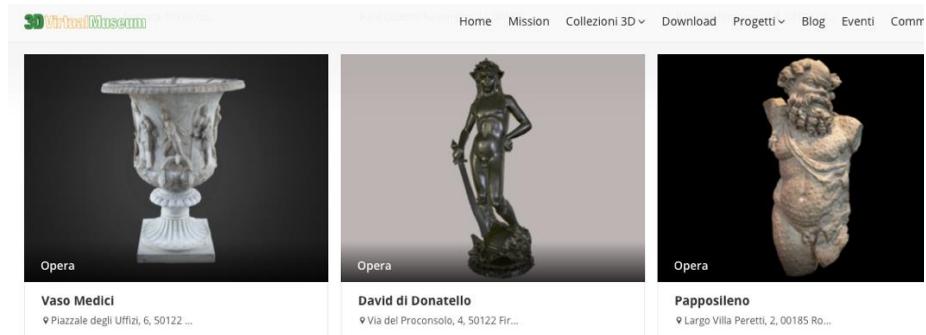


Figure 1.9. 3D models available on the repositories www.3d-virtualmuseum.it

The increasing demand for photogrammetric and laser scanning acquisitions brought in parallel a growing availability of not-interpreted point clouds and 3D models. In this context, the association of semantic information to the point cloud or mesh is beneficial to enrich the data. Considering the consistency of some dataset, it becomes fundamental overcoming the time-consuming manual procedure of classification, introducing automated segmentation and classification methods. The former refers to group points in subsets (commonly called segments) characterised by having one or more properties in common (geometric, radiometric, etc.) whereas classification means the definition and assignment of points to specific classes ("labels") according to different criteria. In the next chapter will be described the aim of the research, related to the identification of automated procedure for 3d heritage classification.

Part 1. Introduction and State of the art

REFERENCES

- Abate, D., Menna, F., Remondino, F. and Gattari, M.G., 2014. 3D painting documentation: evaluation of conservation conditions with 3D imaging and ranging techniques. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 45.
- Apollonio, F.I., Basilissi, V., Callieri, M., Dellepiane, M., Gaiani, M., Ponchio, F., Rizzo, F., Rubino, A.R. and Scopigno, R., 2018. A 3D-centered information system for the documentation of a complex restoration intervention. *Journal of Cultural Heritage*, 29, pp.89-99.
- Barsanti, S.G., Remondino, F., Fenández-Palacios, B.J. and Visintini, D., 2014. Critical factors and guidelines for 3D surveying and modelling in Cultural Heritage. *International Journal of Heritage in the Digital Era*, 3(1), pp.141-158.
- Baik, A.H.A., Yaagoubi, R. and Boehm, J., 2015. Integration of Jeddah historical BIM and 3D GIS for documentation and restoration of historical monument. International Society for Photogrammetry and Remote Sensing (ISPRS).
- Bekele, M.K., Pierdicca, R., Frontoni, E., Malinvern, E.S. and Gain, J., 2018. A survey of augmented, virtual, and mixed reality for cultural heritage. *Journal on Computing and Cultural Heritage (JOCCH)*, 11(2), p.7.
- Callieri, M., Pingi, P., Potenziani, M., Dellepiane, M., Pavoni, G., Lureau, A. and Scopigno, R., 2015. Alchemy in 3D: A digitization for a journey through matter. In *2015 Digital Heritage*, Vol. 1, pp. 223-230. IEEE.
- Chiabrando, F., Sammartano, G., Spanò, A. and Semeraro, G., 2017. Multi-temporal images and 3D dense models for archaeological site monitoring in Hierapolis of Phrygia (TK). *Archeologia e Calcolatori*, 28(2), pp.469-484.
- 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.
- Fernández-Palacios, B.J., Morabito, D. and Remondino, F., 2017. Access to complex reality-based 3D models using virtual reality solutions. *Journal of cultural heritage*, 23, pp.40-48.
- Gomes, L., Bellon, O.R.P. and Silva, L., 2014. 3D reconstruction methods for digital preservation of cultural heritage: A survey. *Pattern Recognition Letters*, 50, pp.3-14.
- Gruen, A., 2008. Reality-based generation of virtual environments for digital earth. *Int. Journal of Digital Earth*, 1(1), pp. 88-106.
- Guidi, G., Russo, M. and Angheloddu, D., 2014. 3D survey and virtual reconstruction of archeological sites. *Digital Applications in Archaeology and Cultural Heritage*, 1(2), pp.55-69.
- Guidi, G., Barsanti, S.G., Micoli, L.L. and Russo, M., 2015. Massive 3D digitization of museum contents. In *Built heritage: Monitoring conservation management* (pp. 335-346). Springer, Cham.
- Lowenthal, D., 2015. *The past is a foreign country-revisited*. Cambridge University Press.
- Menna, F., Agrafiotis, P. and Georgopoulos, A., 2018. State of the art and applications in archaeological underwater 3D recording and mapping. *Journal of Cultural Heritage*, 33, pp.231-248.
- Remondino, F., 2011. Heritage recording and 3D modeling with photogrammetry and 3D scanning. *Remote sensing*, 3(6), pp.1104-1138.

Chapter 1. 3D technologies for Cultural Heritage

Remondino, F., Rizzi, A., Jimenez, B., Agugiaro, G., Baratti, G. and De Amicis, R., 2011. The Etruscans in 3D: From space to underground. *Geoinformatics FCE CTU*, 6, pp.283-290.

Roussel R., Bagnérès M., De Luca L. and Bomblet P., 2019. A digital diagnosis for the <<autumn>> statue (Marseille, France): photogrammetry, digital cartography and construction of a thesaurus. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XLII-2/W15.

CHAPTER 2

An overview of segmentation and classification methods

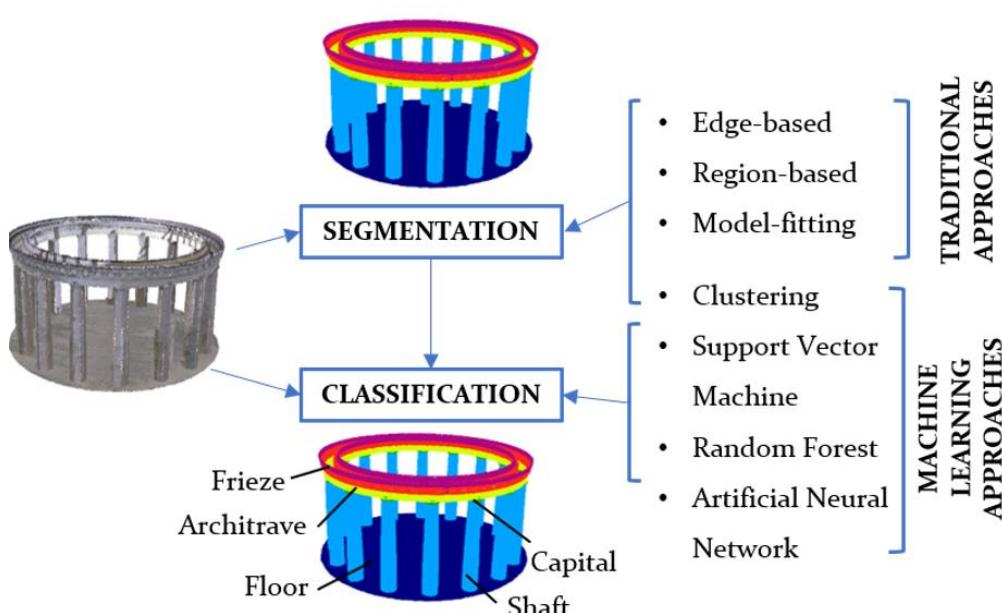


Figure 2.1. Synthetic representation of the segmentation and classification methods.

Part 1. Introduction and State of the art

Most 3D segmentation and classification methods have some root to image segmentation/classification. Segmentation is the process of grouping data (e.g., images, point clouds, or meshes) into multiple homogeneous regions with similar properties (Grilli et al., 2017). These regions are homogeneous regarding some criteria, called features, that constitute a characteristic property or set of properties that are unique, measurable, and differentiable. For 2D imagery, features refer to visual properties such as size, colour, shape, scale patterns, etc. Concerning point clouds data, features are typically connected with specific geometric characteristics of the local distribution in the neighbourhood of the points (Weinmann, 2016) such as surface normals, gradients, curvature (Further details in Chapter 5).

Once 2D or 3D scenarios have been segmented, each group can be classified with a label/ class. Classification procedure gives the parts/segments some semantics; hence, classification is often called semantic segmentation.

The concept behind segmentation and classification is to enrich the representation of an image/point cloud/mesh into something more meaningful and easier to analyse. Both 2D and 3D classification are fundamental tasks in various application, such as geospatial environment (Pal 2005), object detection (Cheng and Han, 2016), medical analyses (Shen et al., 2017), license plate and vehicle recognition (Li et al., 2018), classification of microorganisms (Li et al., 2015), fruit recognition (Dubey et al., 2013) and much more (Llamas et al. 2017).

The author in the following sections presents a literature overview of the topic. The traditional segmentation methods (edge-based, region-based and model fitting) are separated from the Machine Learning approaches (Clustering, Support Vector Machine (SVM), Random Forest (RF), Deep Neural Network) (Figure 2.1).

2.1 TRADITIONAL SEGMENTATION APPROACHES

2.1.1 Edge-based segmentation

Image segmentation is the process of partitioning an image into parts or regions (Yuheng and Hao, 2017). This division into parts is often based on the characteristics of the pixels in the image. One way to find regions in an image is to look for abrupt discontinuities in pixel values, which typically indicate edges. Therefore, these edges can define regions/segments (Al-Amri et al., 2010; Arbelaez et al., 2010; Kaur et al., 2012). Similarly, 3D edge-based segmentation algorithms are composed of two different steps:

- Edge detection: the borders of different regions are outlined where changes in the local surface properties of the points exceed a given threshold. The properties most commonly used are surface normals, gradients, principal curvatures, and higher-order derivatives.

Chapter 2. An overview of segmentation and classification methods

- Grouping: the points inside the boundaries are grouped, resulting in the final segments.

Conventional variations have been described in the literature. Sappa and Devy (2001) propose a fast segmentation strategy based on the extraction of closed contours from a binary edge map. Wang and Shan (2009) segment LiDAR point clouds for extracting building boundaries using a local convex hull approach to detect the edges. In Ni et al. (2016), an automated method is presented, for detecting 3D edges and tracing feature lines from 3D-point clouds.

Even if these methods can perform fast segmentations, they may produce not accurate results as they strictly depend on the noise or sparse distribution of the point clouds. Moreover, the detection of disconnected edges can make the identification of the regions difficult without a filling or interpretation procedure (Castillo et al., 2013).

2.1.2 Region-based segmentation

Region-based segmentation methods work with region-growing algorithms and can be divided into:

- *Bottom-up approaches*: they start from some seed points and grow the segments based on given similarity criteria. Seeded region approaches are highly dependent on selected seed points. Inaccurate selection of seed points can affect the segmentation process and cause under- or over-segmentation results.
- Top-down approaches: they start by assigning all points to one group, then it subdivides the entire region into smaller ones. Where and how to subdivide unseeded regions remains the main difficulty of these methods that require prior knowledge (e.g., object models, number of regions, etc.) (Nguyen, A. and Le, B., 2013).

In 2D, these methods divide the image into regions based on colour values or similar intensity-based rules, starting from one or more points (*seed points*).

When it comes to point clouds segmentation, again starting from the seed points, the regions grow around neighbouring points with similar characteristics. The regions are commonly defined by geometric similarities, such as surface orientation, curvature, etc. (Rabbani et al., 2006; Jagannathan and Miller, 2007).

The initial algorithm was introduced by Besl et al. (1988), and then several variations were presented in the literature. The region growing method proposed by Vosselman et al. (2004) has introduced the use of colour properties beside geometrical criteria. The surface normal and curvatures constraints were widely used to find the smoothly connected areas (Klasing et al., 2009; Belton and Lichti, 2006) whereas Xiao et al. (2013) proposed to use sub-window as the growth unit. Vo et al. (2015) presented an octree-based region growing approach for a fast surface patch segmentation of urban environment 3D point clouds. Recently Che and Olsen (2017) have successfully combined edge- and region-

Part 1. Introduction and State of the art

based methods to perform segmentation on terrestrial laser scanner data (Figure 2.2).

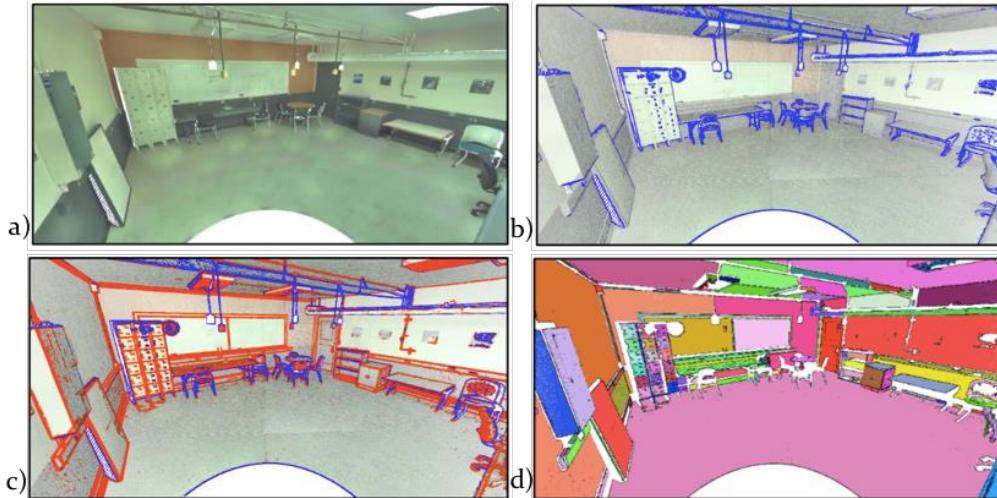


Figure 2.2. Segmentation results based on edge detection: the initial TLS data coloured by image texture (a); points lying on silhouettes (blue) are extracted (b); points with large variation in normals are extracted as edges (orange) (c); grouping of smooth surfaces following region growing (d)(Che and Olsen, 2017).

A collection of region growing algorithms is available in the Point Cloud Library (Rusu and Cousins, 2011). Figure 2.3 shows the results of a segmentation approach completed by a region growing algorithm implemented in the PCL. The purpose of the algorithm is to merge/join similar points and deliver a set of clusters with points belonging to the same smooth surface.

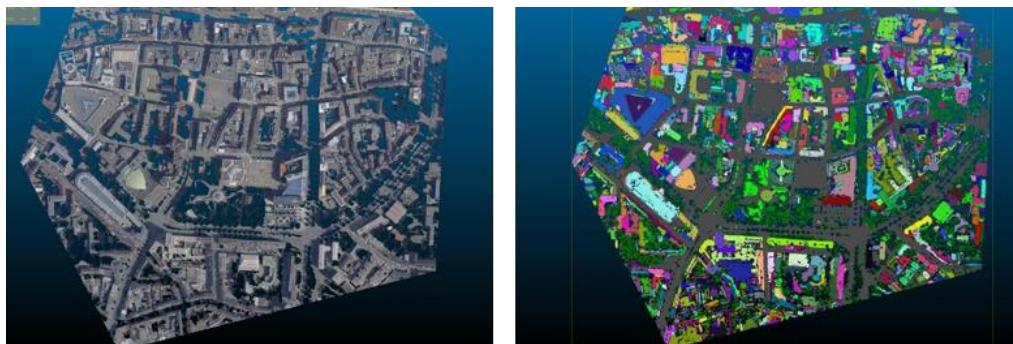


Figure 2.3. Point cloud segmented with a region growing algorithm available in the Point Cloud Library (PCL)(Özdemir and Remondino, 2018) .

In general, the region growing methods are more robust to noise than the edge-based ones because of the using of global information (Liu and Xiong, 2008). However, these methods are sensitive to (i) the location of initial seed regions and (ii) inaccurate estimations of the normals and curvatures of points near region boundaries.

2.1.3 Segmentation by model fitting

This approach is based on the observation that many human-made objects can be decomposed into geometric primitives like planes, cylinders, and spheres.

Chapter 2. An overview of segmentation and classification methods

Therefore, primitive shapes are fitted onto point cloud data, and the points that conform to the mathematical representation of the primitive shape are labelled as one segment (Figure 2.4). In case the primitives have some semantic meaning, such an approach is also performing a classification. As part of the model fitting-based category, two widely employed algorithms are the Hough Transform (HT) (Ballard, 1981) and the Random Sample Consensus (RANSAC) approach (Fischer and Bolles, 1981).

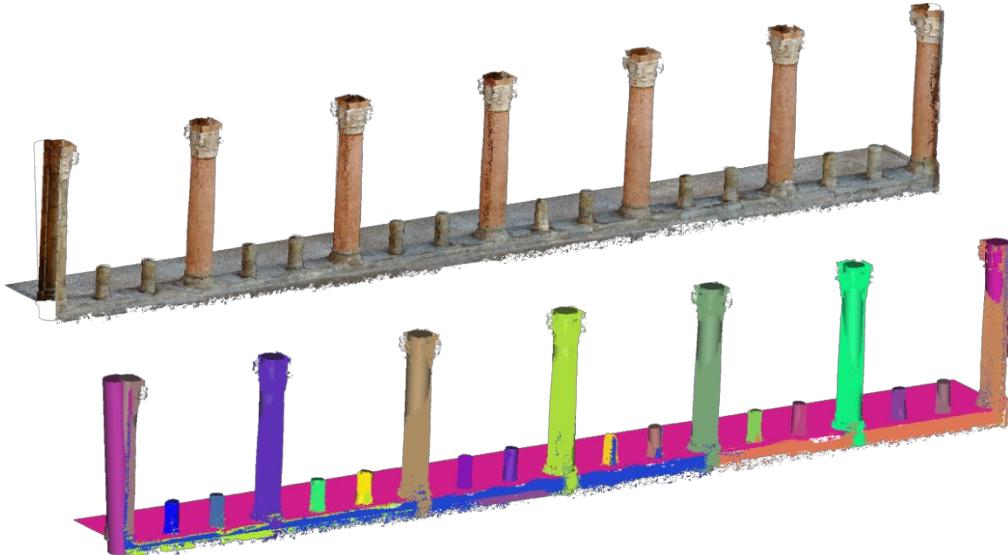


Figure 2.4. Segmentation of a 3D point cloud by primitive geometric fitting. Dataset of the porticoes of Bologna (Remondino et al., 2016).

The HT has been used to detect planes (Vosselman et al., 2004), cylinders, and spheres (Rabbani et al., 2006). The RANSAC method is used to extract shapes by randomly drawing minimal data points to construct candidate shape primitives. The candidate shapes are checked against all points in the dataset to determine a value for the number of points that represents the best fit. Tarsha-Kurdi et al. (2007) compared RANSAC and 3D HT for automatically detecting roof planes from LiDAR-based point clouds. Despite the limitation encountered in both methods, RANSAC is considered more efficient concerning segmented results and running time. It can process a large amount of input data in negligible time. On the other hand, 3D HT is slower and more sensitive to the segmentation parameters values. Chen et al. (2014) have proposed a modified RANSAC segmentation algorithm less sensitive to noise, that maintains topological consistency and avoids over and under-segmentation of building primitives.

Model fitting methods are fast and robust with outliers (Poux et al., 2016). Their efficiency for the 3D detection of simple shapes such as cylinders, spheres, cones, torus, planes, and cubes has been proven. However, in the architectural field, details cannot always be modelled into easily recognisable geometrical shapes. Thus, if geometric properties can characterise some entities, others are more readily distinguished by their colour content (Barnea and Filin, 2013).

2.2 MACHINE LEARNING CLASSIFICATION APPROACHES

Different standard classification approaches have been proposed in the literature (Section 2.1). However, only recently, significant progress has come out in automatic procedures for 2D and 3D classification, thanks to the advent of the Artificial Intelligence (AI) (Noh et al., 2015; Badrinarayanan et al., 2017) and (Weinmann et al., 2014; Guo et al., 2015; Hackel et al., 2017; Qi et al., 2017a-b; Grilli et al., 2019a-b), respectively.

Machine and Deep Learning (ML / DL) are fields of AI scientific research concerned with the development of algorithms that allow machines to make a prediction based on empirical training data. Associated with the training data are the features, variables found in the given training set that can powerfully/sufficiently help us build an accurate predictive model. To give a practical example, for a predictive model which have to classify types of cat, features could be size, colour, shape, scale patterns of the cat. After learning the correlation between these features and the class of cat, the model becomes able to determine which class of cat a given instance is, only by looking at his features.

Machine learning algorithms can be divided into two paradigms according to the learning approach followed:

- *Supervised learning algorithms*: they learn from both the data features and the labels associated with which. The trained model is then used to provide a semantic classification of the entire dataset (Figure 2.5). If for the methods mentioned above (Section 2.1), the classification is a step after the segmentation, when using supervised machine learning methods, the class labelling procedure is planned before to segment the model. Different standard algorithms such as Support Vector Machines (SVM), Decision trees and Random Forest (RF), and artificial neural networks are described below.

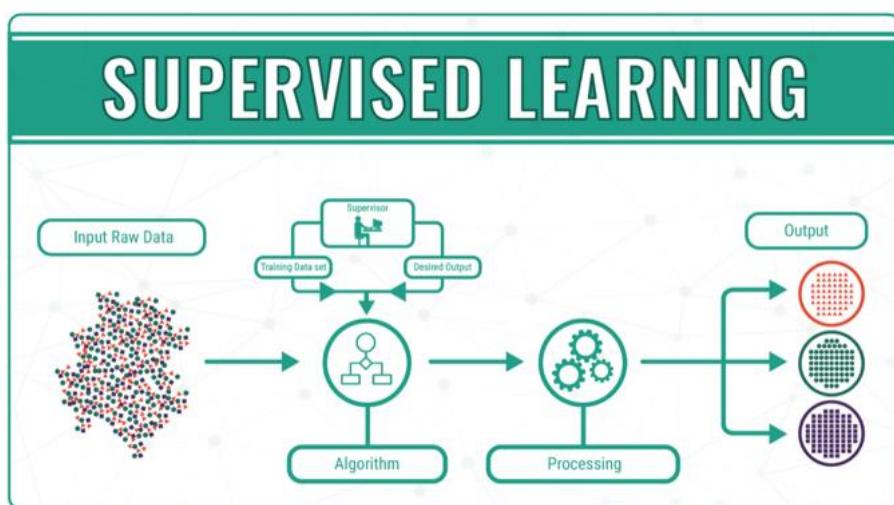


Figure 2.5. Supervised machine learning workflow (Thapliyal, 2019).

- *Unsupervised learning algorithms*: they take the features of data points without the need for labels, as the algorithms introduce their own enumerated labels. No annotations are requested, but the outcome might not be aligned with the user's intention (Figure 2.6). The most widely used unsupervised approach for classification problems is clustering, and notably, the K-means algorithm is broadly used for its simplicity of implementation and convergence speed (Section 2.2.3).

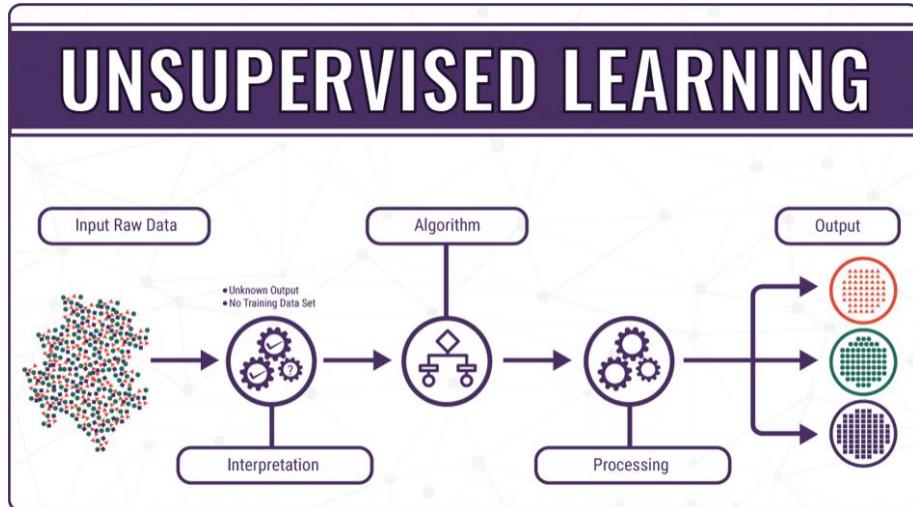


Figure 2.6. Unsupervised machine learning workflow (Thapliyal, 2019).

2.2.1 Support Vector Machine

A Support Vector Machine (SVM) is a binary classifier formally defined by a separating hyperplane (Scholkopf and Smola 2001). In the simplest of terms, once trained with labelled data SVMs use a linear model where they can separate the data, by identifying the “best” line that separates the data; this separation forms the classification (Pedregosa, et al., 2011). Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane (Figure 2.7).

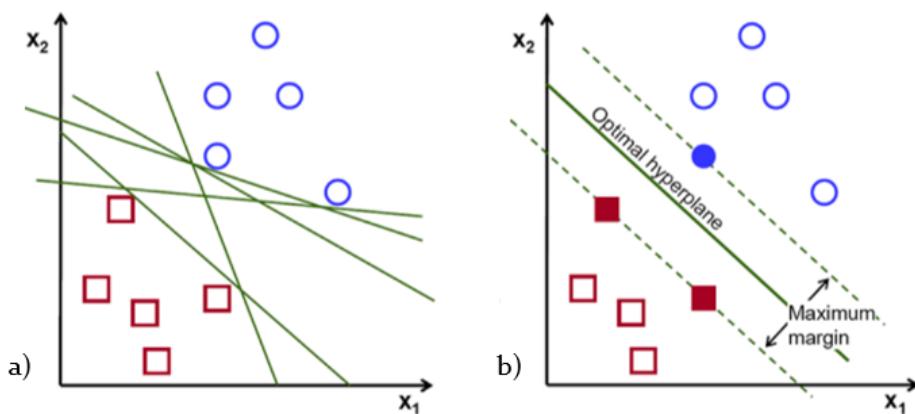


Figure 2.7. Possible hyperplanes (**a**); hyperplane with maximum margin (**b**) (Singh Chaun, 2019)

Part 1. Introduction and State of the art

The applications of SVMs are extensive including, but not limited to; Face Detection, Image Classification, Hand-writing recognition, and Bioinformatics. SVMs can generalise well on difficult image classification problems (Chapelle et al., 1999) and have shown promising results and developments in image classification specifically in remote sensing fields (Pal and Mather, 2005). The most common application within remote sensing are land cover/use, and many algorithms have been developed as part of the SVM group (Kavzoglu and Colkesen, 2009) to provide information about the Earth's surface in many wide-scale applications. In more recent years, SVMs models have been applied to the classification of point clouds (Bogdan and Cousins, 2011; Caputo et al., 2015). The binary classification has been investigated to extract tree species (Bohm et al., 2016), road surfaces (Shu et al., 2016), land cover (Zhou et al., 2016), and construction sites (Xu et al., 2016).

SVMs are highly effective at classifying small and medium datasets, while they are not very efficient computationally if the dataset is vast (Romero et al., 2015).

2.2.2 Decision trees and Random Forest

Random Forest (Breiman, L., 2001), is one of the most used supervised learning algorithms for classification problems (Bosh et al., 2007; Rodriguez-Galiano et al., 2012). During the training phase, both the features and the labels are given as input to the model so it can learn to classify points based on the features. It uses an ensemble of classification trees, gets a prediction from each tree, and selects the best solution through voting (Figure 2.8). It is possible to think of a decision tree as a series of yes/no questions asked about our data, eventually leading to a predicted class. The decision tree tries to form nodes containing a high proportion of samples (data points) from a single class by finding values in the features that cleanly divide the data into classes. While great for producing models that are easy to understand and implement, decision trees also tend to overfit on their training data—making them perform poorly if data shown to them later don't closely match to what they were trained on. Overfitting occurs when we have a very flexible model (the model has a high capacity) which essentially memorises the training data by fitting it closely. The problem is that the model learns not only the actual relationships in the training data but also any noise that is present. A flexible model is said to have high variance because the learned parameters (such as the structure of the decision tree) vary considerably with the training data. Starting from these considerations, RFs rather than just merely averaging the prediction of trees uses two key concepts to avoid the overfitting problems:

- bootstrap = Random subsets of the considered features when splitting nodes;
- bagging= Random sampling of training data points when building trees.

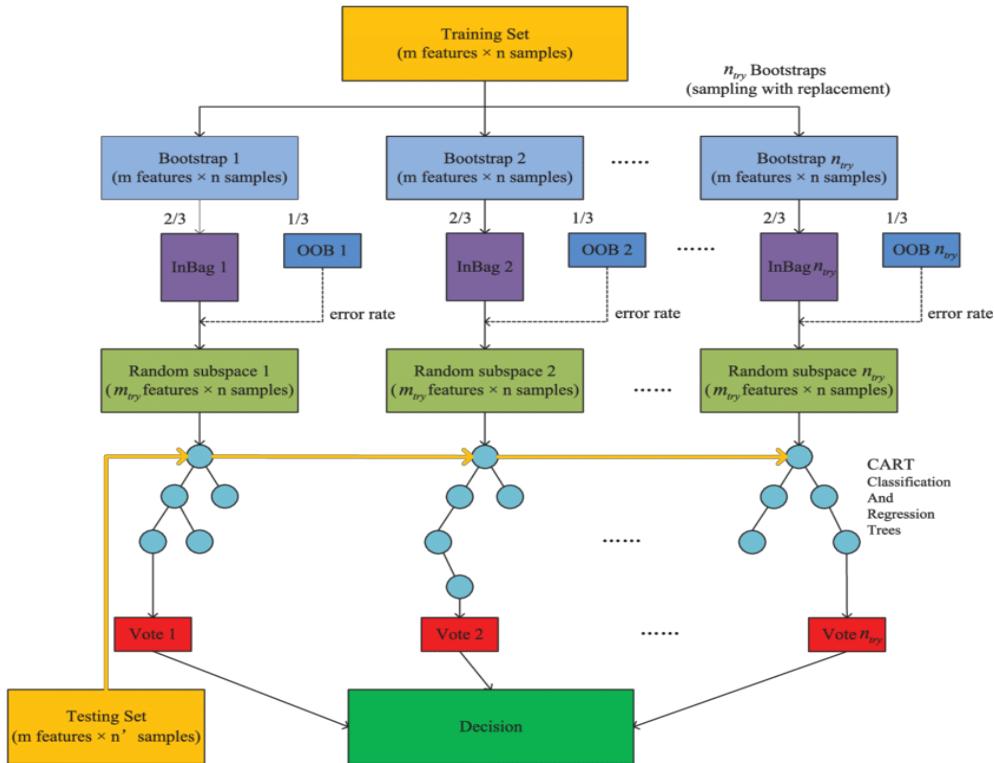


Figure 2.8. The workflow of the Random Forest algorithm (Han et al., 2018).

RF has been used extensively in point cloud classification (Chehata et al., 2009; Niemeyer et al., 2014; Weinmann et al., 2015; Hackel et al., 2016); a detailed review of the different applications in Remote Sensing is given in Belgiu and Drăgut (2016). According to Weinmann et al.'s extensive work (Weinmann et al., 2013; Weinmann et al., 2015; Weinmann, 2016), RF can be considered as one of the most suitable classifiers for point cloud analysis.

With RF, it is possible to train a model with a relatively small number of samples and get pretty good results. However, it quickly reaches a point where more samples do not improve the accuracy. In contrast, a deep neural network (Section 2.2.4) needs more samples to deliver the same level of accuracy, but it benefits from massive amounts of data, and continuously improve the accuracy.

2.2.3 K-means clustering

Clustering is a type of unsupervised machine learning that segments similar data points into groups, called clusters. Objects in the same clusters are more like each other than those in other groups; the clusters should have minimal variance. Clustering algorithms should be applied to the nature of the problem considering the characteristics of the objects in question (Naik and Shah, 2014).

K-Means is a clustering algorithm that divides observations into k clusters using features, where each observation belongs to the cluster with the nearest mean. One important detail about K-Means Clustering is that, even though it identifies which data point should part of which cluster, the operator has to specify the parameter K , representing the total number of clusters that he wants to use to "distribute" your data.

Part 1. Introduction and State of the art

The original K-means algorithm presented by MacQueen (1967) has been widely exploited by various researchers for image (Chitade and Katiyar, 2010; Saraswathi and Allirani, 2013) and point clouds (Teutsch et al., 2011; Zhang et al., 2015; Zhang et al., 2016).

K-means algorithm are easy to implement and good at segmenting large datasets. However, they can have trouble at clustering data where clusters are of varying sizes and density (Raykov et al., 2016).

2.2.4 Artificial neural networks

Among the different tools used in machine learning, there are also artificial neural networks. As suggested by the name, they are systems which intend to replicate the way humans learn. A neural network consists of an input layer, a hidden layer(s) and an output layer. Every layer consists of nodes, loosely modelled from neurons in the brain (Figure 2.9).

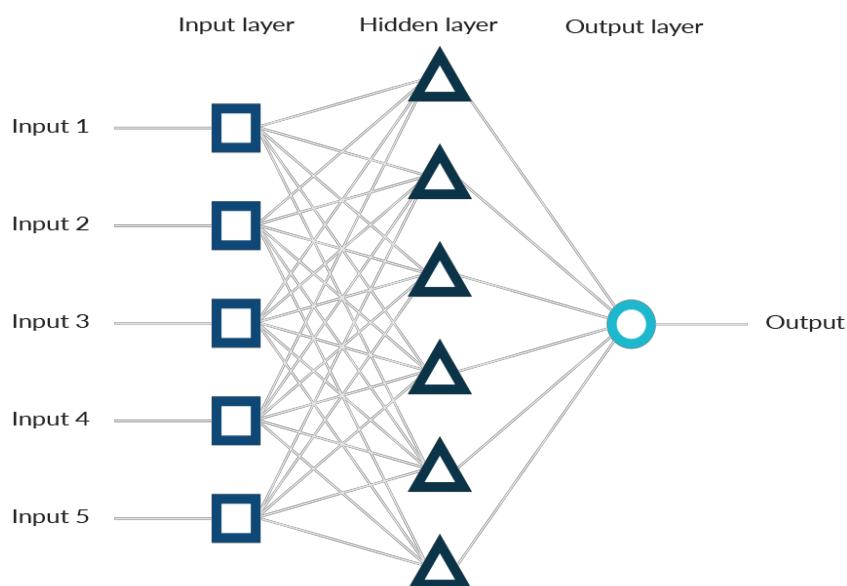


Figure 2.9. Structure of an Artificial Neural Network (Lofwander, S., 2017).

While neural networks have been around since a while (McCulloch and Pitts, 1943), they have become a significant part of artificial intelligence only in the last several decades (Fukushima, 1980). A significant advance has been the arrival of deep learning neural networks (LeCun et al., 2015). The difference between neural networks and deep learning lies in the depth of the model. The term deep learning refers to complex neural networks, in which the different layers extract different features until they can recognise what they are looking for.

Deep learning models can be considered an evolution of the machine learning ones, as they can learn by themselves the features, as part of the training process. On the other side, in ML is the operator that identifies the features that are necessary for the classification process. This ability to learn features is often seen as the cause for the rapid improvement in 2D and 3D understanding benchmark results.

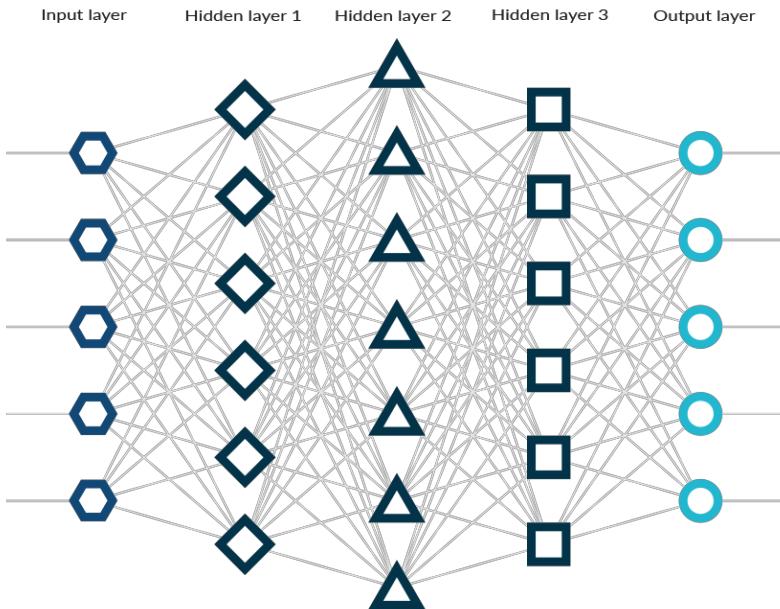


Figure 2.10. Structure of a Deep Neural Network (Lofwander, S., 2017).

Deep neural networks have already shown impressive performance on regularly structured data representations such as images and time series (Liu et al., 2017; Garcia-Garcia et al., 2018). Nevertheless, promising results on supervised learning tasks such as object classification and semantic segmentation, operating on raw point cloud data, have been presented (Qi et al., 2017a-b; Wang et al., 2018; Zhou et al., 2018). A complete review of the state-of-the-art deep learning methods for classification, object detection and point-wise segmentation of 3D sensed data was recently published by Griffith and Boehm (2019).

2.3 CLASSIFICATION STRATEGIES APPLIED TO CULTURAL HERITAGE

In the field of cultural heritage, processes such as segmentation and classification can be applied at different scales, from entire archaeological sites and landscapes to small artefacts.

Different solutions for the classification of architectural images were presented in the literature. Among them, pattern detection (Chu et al., 2012), Gabor filters and support vector machine (Mathias et al., 2012), K-means algorithms (Oses et al., 2014), clustering and learning of local features (Shalunts et al., 2011), hierarchical sparse coding of blocks (Zhang et al., 2014).

While the use of machine learning techniques for point cloud classification has been successfully investigated in the geospatial environment (Niemeyer et al., 2014; Guo et al., 2015; Weinmann et al., 2014; Qi et al., 2017a-b; Özdemir and Remondino, 2019), only recently has been explored in Cultural Heritage (CH) application (Poux et al., 2017; Grilli et al., 2018; Grilli et al., 2019a-b; Malinverni et al., 2019). Several benchmarks have been proposed in the Geomatics

Part 1. Introduction and State of the art

community, providing labelled terrestrial and airborne data on which users can test and validate their algorithms. Most of the available datasets provide classified natural, urban, and street scenes such as Semantic3D (Hackel et al., 2017) or The Cityscapes Dataset (Cordts et al., 2016). While in those scenarios, the object classes and labels are almost defined (mainly ground, roads, trees, and buildings), the identification of precise categories in the heritage field is much more complicated:

- for the same case study several classes can be identified based upon different purposes;
- not always a semantic architectural class is linked to a precise shape/colour.

Probably for these reasons, up to now the only available databases of annotated heritage are in 2D images and refer only to building facades such as the Ecole Centrale Paris (ECP) Facades dataset (Teboul et al., 2010), eTRIMS (Korc and Forstner, 2009), and CMP Facade Database (Tyleček and Šára, 2013). Despite this existing-data shortage, different machine learning approaches were proposed in the 2D architectural and heritage context. Osés et al. (2014) have used different machine learning classifiers to perform an image-based delineation of masonry walls. Amato et al. (2015) use k-nearest neighbour (kNN) classification and landmark recognition techniques to address the problem of monument recognition in images. Convolutional Neural Networks (CNN) was applied for the first time to heritage scenarios in Llamas et al. (2016) and Llamas et al., (2017). CNNs are also used by Yasser et al. (2017) for visual categorisation. The author also proposes to develop a digital heritage search platform (ICARE) that allows users to archive digital heritage content and perform semantic queries over multimodal cultural heritage data archives.

Regarding the 3D dimension, in most of the literature available classification processes are performed for annotation and restoration purposes, and the information are transferred from 2D to 3D (Campanaro et al., 2016; Grilli et al., 2018). For instance, the web platform Aioli gives a similar example of annotation onto 3D heritage (e.g., www.aioli.cloud), where the 2D mapping data is in real-time displayed onto the 3D model (Roussel et al., 2019). The developers of the platform are willing to integrate some ancillary tools, based on supervised methods, to make the processes of annotation semi-automatic.

Concluding, there is a rich literature about classification methods applied to architectural images, but only a few references refer to automatic classification strategies for 3D architectural heritage.

REFERENCES

- Al-Amri, S.S. and Kalyankar, N.V., 2010. Image segmentation by using threshold techniques. *arXiv preprint arXiv:1005.4020*.
- Amato, G., Falchi, F. and Gennaro, C., 2015. Fast Image Classification for Monument Recognition. *Journal on Computing and Cultural Heritage*, 8, Volume 8, pp. 1-25.
- Arbelaez, P., Maire, M., Fowlkes, C. and Malik, J., 2010. Contour detection and hierarchical image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 33(5), pp.898-916.
- Badrinarayanan, V., Kendall, A. and Cipolla, R., 2017. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12, Volume 39, pp. 2481-2495.
- Ballard, D. H., 1981. Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition*, 1, Volume 13, pp. 111-122.
- Barnea, S. and Filin, S., 2013. Segmentation of terrestrial laser scanning data using geometry and image information. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2, Volume 76, pp. 33-48.
- Belgiu, M. and Drăguț, L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 4, Volume 114, pp. 24-31.
- Belton, D. and Lichti, D.D., 2006. Classification and segmentation of terrestrial laser scanner point clouds using local variance information. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(5), pp.44-49.
- Besl, P. J. and Jain, R. C., 1988. Segmentation through variable-order surface fitting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3, Volume 10, pp. 167-192.
- Bogdan, R. R. and Cousins, S., 2011. 3D is here: Point Cloud Library (PCL). Shanghai, IEEE International Conference on Robotics and Automation (ICRA).
- Böhm, J. et al., 2016. The Iqmulus urban showcase: Automatic tree classification and identification in huge mobile mapping point clouds. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 6, Volume XLI-B3, pp. 301-307.
- Bosch, A., Zisserman, A. and Munoz, X., 2007. Image Classification using Random Forests and Ferns. s.l., IEEE.
- Breiman, L., 2001. Random Forests. *Machine Learning*, Volume 45, pp. 5-32.
- Campanaro, D. M., Landeschi, G., Dell'Unto, N. and Touati, A.-M. L., 2016. 3D GIS for cultural heritage restoration: A 'white box' workflow. *Journal of Cultural Heritage*, 3, Volume 18, pp. 321-332.
- Caputo, M., Denker, K. and Franz, M. O., 2015. Support Vector Machines for Classification of Geometric Primitives in Point Clouds. Paris, 8th International Conference on Curves and Surfaces.
- Castillo, E., Liang, J. and Zhao, H., 2013. Point Cloud Segmentation and Denoising via Constrained Nonlinear Least Squares Normal Estimates. In *Innovations for Shape Analysis*. Springer Berlin Heidelberg, pp. 283-299.
- Chapelle, O., Haffner, P. and Vapnik, V. N., 1999. Support Vector Machines for Histogram-Based Image Classification. *IEEE Transactions on Neural Networks*, 10(5), pp. 1055-1064.
- Che, E. and Olsen, M. J., 2017. Fast Edge Detection for Terrestrial Laser Scanning through Normal Variation Analysis. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 9, Volume IV-2/W4, pp. 51-57.

Part 1. Introduction and State of the art

- Chehata, N., Guo, L. and Mallet, C., 2009. Airborne lidar feature selection for urban classification using random forests. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 38 (Part 3).
- Chen, D., Zhang, L., Mathiopoulos, P. T. and Huang, X., 2014. A Methodology for Automated Segmentation and Reconstruction of Urban 3-D Buildings from ALS Point Clouds. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10, Volume 7, pp. 4199-4217.
- Cheng, G. and Han, J., 2016. A survey on object detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 7, Volume 117, pp. 11-28.
- Chitade, A. Z. and Katiyar, S. k., 2010. Color based image segmentation using K-means clustering. *International Journal of Engineering Science and Technology*, 2(10), pp. 5319-5325.
- Chu, W.T. and Tsai, M.H., 2012. Visual pattern discovery for architecture image classification and product image search. In *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval*, p. 27. ACM.
- Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S. and Schiele, B., 2016. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3213-3223.
- Dubey, S. R., Dixit, P., Singh, N. and Gupta, J. P., 2013. Infected Fruit Part Detection using K-Means Clustering Segmentation Technique. *International Journal of Interactive Multimedia and Artificial Intelligence*, Volume 2, p. 65.
- Fischler, M. A. and Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 6, Volume 24, pp. 381-395.
- Fukushima, K., 1980. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 4, Volume 36, pp. 193-202.
- Garcia-Garcia, A., Orts-Escalano, S., Oprea, S., Villena-Martinez, V., Martinez-Gonzalez, P. and Garcia-Rodriguez, J., 2018. A survey on deep learning techniques for image and video semantic segmentation. *Applied Soft Computing*, 70, pp. 41-65.
- Griffiths, D. and Boehm, J., 2019. A Review on deep learning techniques for 3D sensed data classification. *Remote Sensing*, 11(12), p.1499.
- Grilli, E., Menna, F. and Remondino, F., 2017. A Review of Point Clouds Segmentation and Classification Algorithms. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2, Volume XLII-2/W3, pp. 339-344.
- Grilli, E., Diniinno, D., Petrucci, G. and Remondino, F., 2018. From 2D to 3D Supervised Segmentation and Classification for Cultural Heritage Applications. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 5, Volume XLII-2, pp. 399-406.
- Grilli, E., M. Farella, E., Torresani, A. and Remondino, F., 2019a. Geometric features analysis for the classification of cultural heritage point clouds. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XLII-2/W15, pp. 541-548.
- Grilli, E., Özdemir, E. and Remondino, F., 2019b. Application of machine and deep learning strategies for the classification of heritage point clouds. In proceeding of *The international ISPRS Geospatial Conference 2019*, University of Tehran. (In press)
- Guo, B., Huang, X., Zhang, F. and Sohn, G., 2015. Classification of airborne laser scanning data using JointBoost. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2, Volume 100, pp. 71-83.

Chapter 2. An overview of segmentation and classification methods

- Hackel, T., Wegner, J. D. and Schindler, K., 2016. Fast Semantic Segmentation of 3D Point Clouds with Strongly Varying Density. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 6, Volume III-3, pp. 177-184.
- Hackel, T., Savinov, N., Ladicky, L., Wegner, J.D., Schindler, K. and Pollefeys, M., 2017. Semantic3d. net: A new large-scale point cloud classification benchmark. *arXiv preprint arXiv:1704.03847*.
- Han, T., Jiang, D., Zhao, Q., Wang, L. and Yin, K., 2018. Comparison of random forest, artificial neural networks and support vector machine for intelligent diagnosis of rotating machinery. *Transactions of the Institute of Measurement and Control*, 40(8), pp.2681-2693.
- Jagannathan, A. and Miller, E. L., 2007. Three-Dimensional Surface Mesh Segmentation Using Curvedness-Based Region Growing Approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12, Volume 29, pp. 2195-2204.
- Kaur, J., Agrawal, S. and Vig, R., 2012. A Comparative Analysis of Thresholding and Edge Detection Segmentation Techniques. *International Journal of Computer Applications*, 2, Volume 39, pp. 29-34.
- Kavzoglu, T. and Colkesen, I., 2009. A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation*, 11(5), pp. 352-359.
- Klasing, K., Althoff, D., Wollherr, D. and Buss, M., 2009. Comparison of surface normal estimation methods for range sensing applications. In *2009 IEEE International Conference on Robotics and Automation*, pp. 3206-3211. IEEE.
- Korč, F. & Förstner, W., 2009. eTRIMS Image Database for interpreting images of man-made scenes. *Dept. of Photogrammetry, University of Bonn, Tech. Rep. TR-IGG-P-2009-01*.
- LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *nature*, 521(7553), p.436.
- Li, C., Shirahama, K. and Grzegorzek, M., 2015. Application of content-based image analysis to environmental microorganism classification. *Biocybernetics and Biomedical Engineering*, Volume 35, pp. 10-21.
- Li, H., Wang, P., You, M. and Shen, C., 2018. Reading car license plates using deep neural networks. *Image and Vision Computing*, 4, Volume 72, pp. 14-23.
- Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y. and Alsaadi, F.E., 2017. A survey of deep neural network architectures and their applications. *Neurocomputing*, 234, pp.11-26.
- Liu, Y. and Xiong, Y., 2008. Automatic segmentation of unorganized noisy point clouds based on the Gaussian map. *Computer-Aided Design*, 5, Volume 40, pp. 576-594.
- Llamas, J., Lerones, P.M., Zalama, E. and Gómez-García-Bermejo, J., 2016, October. Applying deep learning techniques to cultural heritage images within the INCEPTION project. In *Euro-Mediterranean Conference*, pp. 25-32. Springer, Cham.
- Llamas, J. et al., 2017. Classification of Architectural Heritage Images Using Deep Learning Techniques. *Applied Sciences*, 9, Volume 7, p. 992.
- Lofwander, S., 2017. About Artificial Intelligence, Neural Networks & Deep Learning. [Online] Available at: <https://www.ayima.com/blog/artificial-intelligence-neural-networks-deep-learning.html> [Accessed October 3rd 2019]
- MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1, pp. 281-297.
- Malinverni, E. S., Pierdicca, R., Paolanti, M., Martini, M., Morbidoni, C., Matrone, F., and Lingua, A., 2019. Deep learning for semantic segmentation of 3D point cloud.

Part 1. Introduction and State of the art

International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2/W15, pp. 735–742.

Mathias, M. et al., 2012. Automatic Architectural Style Recognition. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 9, Volume XXXVIII-5/W16, pp. 171-176.

McCulloch, W.S. and Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), pp.115-133.

Naik, D. and Shah, P., 2014. A Review on Image Segmentation Clustering Algorithms. *International Journal of Computer Science and Information Technologies*, 5(3), pp. 3289-3293.

Nguyen, A. and Le, B., 2013, November. 3D point cloud segmentation: A survey. In *2013 6th IEEE conference on robotics, automation and mechatronics (RAM)*, pp. 225-230. IEEE.

Ni, H., Lin, X., Ning, X. and Zhang, J., 2016. Edge Detection and Feature Line Tracing in 3D-Point Clouds by Analyzing Geometric Properties of Neighborhoods. *Remote Sensing*, 9, Volume 8, p. 710.

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*, 1, Volume 87, pp. 152-165.

Noh, H., Hong, S. and Han, B., 2015. Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision*, pp. 1520-1528.

Oses, N., Dornaika, F. and Moujahid, A., 2014. Image-Based Delineation and Classification of Built Heritage Masonry. *Remote Sensing*, 2, Volume 6, pp. 1863-1889.

Özdemir, E. and Remondino, F., 2018. Segmentation of 3D photogrammetric point cloud for 3D building modeling. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XLII-4/W10, pp. 135-142.

Özdemir, E. and Remondino, F., 2019. Classification of Aerial Point Clouds with Deep Learning. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 6, Volume XLII-2/W13, pp. 103-110.

Pal, M. and Mather, P., 2005. Support Vector Machines for Classification in Remote Sensing. *International Journal of Remote Sensing*, 26(5), pp. 1007-1011.

Pedregosa, F. et al., 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, Volume 12, pp. 2825-2830.

Poux, F., Hallot, P., Neuville, R. and Billen, R., 2016. Smart Point Cloud: Definition and Remaining Challenges. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10, Volume IV-2/W1, pp. 119-127.

Poux, F., Neuville, R. and 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*, 8, Volume IV-2/W2, pp. 203-211.

Qi, C.R., Su, H., Mo, K. and Guibas, L.J., 2017a. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 652-660.

Qi, C.R., Yi, L., Su, H. and Guibas, L.J., 2017b. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in neural information processing systems*, pp. 5099-5108.

Chapter 2. An overview of segmentation and classification methods

- Rabbani, T., Van Den Heuvel, F., and Vosselmann, G., 2006. Segmentation of point clouds using smoothness constraint. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 36(5), pp. 248-253.
- Raykov, Y.P., Boukouvalas, A., Baig, F. and Little, M.A., 2016. What to do when K-means clustering fails: a simple yet principled alternative algorithm. *PloS one*, 11(9).
- Remondino, F., Gaiani, M., Apollonio, F., Ballabeni, A., Ballabeni, M. and Morabito, D., 2016. 3D documentation of 40 kilometers of historical porticoes—the challenge. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, p. 41.
- Rodriguez-Galiano, V. F. et al., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 1, Volume 67, pp. 93-104.
- Romero, R., Iglesias, E.L. and Borrajo, L., 2015. A linear-RBF multikernel SVM to classify big text corpora. *BioMed research international*, 2015.
- Roussel R., Bagnéris M., De Luca L. and Bomblet P., 2019. A digital diagnosis for the <>autumn>> statue (Marseille, France): photogrammetry, digital cartography and construction of a thesaurus. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XLII-2/W15, pp. 1039-1046.
- Rusu, R.B. and Cousins, S., 2011, May. Point cloud library (pcl). In *2011 IEEE international conference on robotics and automation*, pp. 1-4.
- Sappa, A.D. and Devy, M., 2001, May. Fast range image segmentation by an edge detection strategy. In *Proceedings Third International Conference on 3-D Digital Imaging and Modeling*, pp. 292-299. IEEE.
- Saraswathi, S. and Allirani, A., 2013. Survey on Image Segmentation via Clustering. *International Journal of Research and Reviews in Information Sciences*, 1(1), pp. 331-335.
- Schölkopf, B. and Smola, A. J., 2018. Learning with Kernels. MIT Press.
- Shalunts, G., Haxhimusa, Y. and Sablatnig, R., 2011, September. Architectural style classification of building facade windows. In *International Symposium on Visual Computing*, pp. 280-289. Springer, Berlin, Heidelberg.
- Shen, D., Wu, G. and Suk, H.-I., 2017. Deep Learning in Medical Image Analysis. *Annual Review of Biomedical Engineering*, 6, Volume 19, pp. 221-248.
- Shu, Z., Sun, K., Qiu, K. and Ding, K., 2016. Pairwise SVM for on-board urban road LiDAR classification. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 6, Volume XLI-B1, pp. 109-113.
- Singh Chauhan, N., 2019. A friendly introduction to Support Vector Machines(SVM). [Online] Available at: <https://towardsdatascience.com/a-friendly-introduction-to-support-vector-machines-svm-925b68c5a079> [Accessed September 7th 2019].
- Tarsha-Kurdi, F., Landes, T. and Grussenmeyer, P., 2007. Joint combination of point cloud and DSM for 3D building reconstruction using airborne laser scanner data. In *2007 Urban Remote Sensing Joint Event*, pp. 1-7. IEEE.
- Teboul, O. et al., 2013. Parsing Facades with Shape Grammars and Reinforcement Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7, Volume 35, pp. 1744-1756.
- Teutsch, C., Trostmann, E. and Berndt, D., 2011. A parallel point cloud clustering algorithm for subset segmentation and outlier detection. In *Videometrics, Range Imaging, and Applications XI*, Vol. 8085, p. 509. International Society for Optics and Photonics.
- Thapliyal, M., 2019. Supervised Learning V/S Unsupervised Learning using Machine Learning. [Online] Available at: <https://medium.com/mlrecipies/supervised-learning->

Part 1. Introduction and State of the art

v-s-unsupervised-learning-using-machine-learning-98de75200415 [Accessed September 7th 2019].

Tyleček, R. and Šára, R., 2013, September. Spatial pattern templates for recognition of objects with regular structure. In *German Conference on Pattern Recognition*, pp. 364-374. Springer, Berlin, Heidelberg.

Vo, A.-V., Truong-Hong, L., Laefer, D. F. and Bertolotto, M., 2015. Octree-based region growing for point cloud segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 6, Volume 104, pp. 88-100.

Vosselman, G., Gorte, B.G., Sithole, G. and Rabbani, T., 2004. Recognising structure in laser scanner point clouds. *International archives of photogrammetry, remote sensing and spatial information sciences*, 46(8), pp.33-38.

Wang, J. and Shan, J., 2009, March. Segmentation of LiDAR point clouds for building extraction. In *American Society for Photogramm. Remote Sens. Annual Conference*, Baltimore, MD, pp. 9-13.

Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M. and Solomon, J.M., 2018. Dynamic graph cnn for learning on point clouds. *arXiv preprint arXiv:1801.07829*.

Weinmann, M., Jutzi, B. and Mallet, C., 2013. Feature relevance assessment for the semantic interpretation of 3D point cloud data. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10, Volume II-5/W2, pp. 313-318.

Weinmann, M., Jutzi, B. and Mallet, C., 2014. Semantic 3D scene interpretation: A framework combining optimal neighborhood size selection with relevant features. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 8, Volume II-3, pp. 181-188.

Weinmann, M., Schmidt, A., Mallet, C., Hinz, S., Rottensteiner, F. and Jutzi, B., 2015. Contextual classification of point cloud data by exploiting individual 3D neighbourhoods. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-3 (2015)*, Nr. W4, 2(W4), pp.271-278.

Weinmann, M., 2016. *Reconstruction and analysis of 3D scenes*. Springer.

Xiao, J. et al., 2013. Three-dimensional point cloud plane segmentation in both structured and unstructured environments. *Robotics and Autonomous Systems*, 12, Volume 61, pp. 1641-1652.

Xu, Y., Tuttas, S., Heogner, L. and Still, U., 2016. Classification of Photogrammetric Point Clouds of Scaffolds for Construction Site Monitoring Using Subspace Clustering and PCA. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 6, Volume XLI-B3, pp. 725-732.

Yasser, A.M., Clawson, K. and Bowerman, C., 2017, July. Saving cultural heritage with digital make-believe: machine learning and digital techniques to the rescue. In *Proceedings of the 31st British Computer Society Human Computer Interaction Conference*, p. 97. BCS Learning & Development Ltd.

Yuheng, S. and Hao, Y., 2017. Image segmentation algorithms overview. *arXiv preprint arXiv:1707.02051*.

Zhang, L., Song, M., Liu, X., Sun, L., Chen, C. and Bu, J., 2014. Recognizing architecture styles by hierarchical sparse coding of blocklets. *Information Sciences*, 254, pp.141-154.

Zhang, K., Bi, W., Zhang, X., Fu, X., Zhou, K. and Zhu, L., 2015. A new k-means clustering algorithm for point cloud. *International Journal of Hybrid Information Technology*, 8(9), pp.157-170.

Zhang, C. and Mao, B., 2016. 3D Building Models Segmentation Based on K-means++ Cluster Analysis. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10, Volume XLII-2/W2, pp. 57-61.

Chapter 2. An overview of segmentation and classification methods

Zhou, M., Li, C. R., Ma, L. and Guan, H. C., 2016. Land Cover Classification from Full-Waveform LIDAR Data Based on Support Vector Machines. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 6, Volume XLI-B3, pp. 447-452.

Zhou, Y. and Tuzel, O., 2018. Voxelnet: End-to-end learning for point cloud based 3d object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4490-4499.

CHAPTER 3

Aim of the research

In recent years, the use of heritage 3D for documentation and dissemination purposes is increasing. If we focus on point clouds, there is a growing need of innovative methods for the treatment and analysis of these data and for their classification, aimed ultimately to exploit in-depth the educational value of these surveys and representations. Point clouds are a powerful collection of geometrical primitives able to represent the shape, size, position, and orientation of objects in space. This information may be augmented with additional contents obtained from other sensors or sources, such as colours, multispectral or thermal information. Due to the complexity and variety of point clouds caused by irregular sampling, varying density, different types of objects, etc., point cloud classification is a very active and challenging research topic. The association of heterogeneous information to 3D data by means of automated classification methods can help to characterise, describe, and better interpret the object under study.

Starting from the existing literature (Chapter 2), the main purpose of the research is to develop, explore and validate reliable and efficient automated procedures for the classification of 3D data (point clouds or polygonal mesh models) of heritage scenarios. Through automatic classification the author aims to:

- distinguish different constructing techniques: it can be useful to deepen the analysis and interpretation of the architecture (Grilli et al., 2018a) (Section 6.1);
- identify existing restoration evidence (Grilli and Remondino, 2018b) (Section 6.4);

Part 1. Introduction and State of the art

- identify and quantify different states of conservation and materials: essential for monitoring or restoration purposes (Grilli and Remondino, 2019) (Section 6.5);
- identify and distinguish structural and decorative architectural elements, highlighting their spatial distribution and organization (Grilli et al., 2019a) (Section 7.3).
- automatically recognize similar architectural elements in vast datasets: propaedeutic for Heritage Building Information Modeling (HBIM) (Grilli et al., 2019b) (Section 7.4 – 7.5).

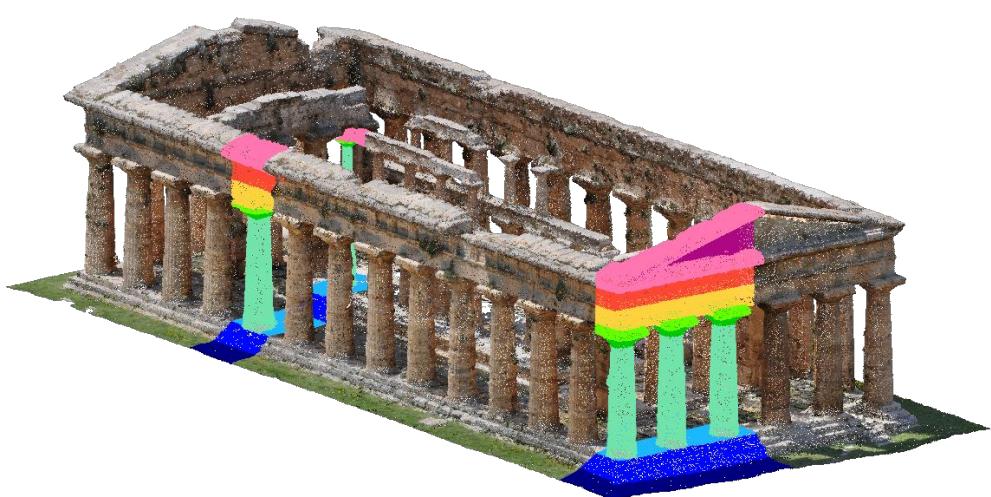
The procedures that are going to be explained are based on Artificial Intelligence methods, and in particular on Machine Learning algorithms (Section 2.2). For the development of the research, various archaeological and architectural scenarios have been taken into consideration, in order to offer a reliable methodology which can be replicated and deployed in various heritage cases (Part 3).

REFERENCES

- Grilli, E., Menna, F. and Remondino, F., 2017. A review of point clouds segmentation and classification algorithms. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, p.339-346.
- Grilli, E., Dininno, D., Petrucci, G. and Remondino, F., 2018a. From 2D to 3D supervised segmentation and classification for cultural heritage applications. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42(2), pp. 399-406.
- Grilli, E., Dininno, D., Marsicano, L., Petrucci, G. and Remondino, F., 2018b. Supervised segmentation of 3D cultural heritage. In *2018 3rd Digital Heritage International Congress (DigitalHERITAGE) held jointly with 2018 24th International Conference on Virtual Systems & Multimedia (VSMM 2018)*, pp. 1-8. IEEE.
- Grilli, E. and Remondino, F., 2019. Classification of 3D Digital Heritage. *Remote Sensing*, 11(7), p.847.
- Grilli, E., M. Farella, E., Torresani, A. and Remondino, F., 2019a. Geometric features analysis for the classification of cultural heritage point clouds. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XLII-2/W15, pp 541-548.
- Grilli, E., Özdemir, E., and Remondino, F., 2019b Application of machine and deep learning strategies for the classification of heritage point clouds. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. XLII-4/W18, pp. 447-454.

PART 2

Proposed methods



CHAPTER 4

Textured-based classification approach

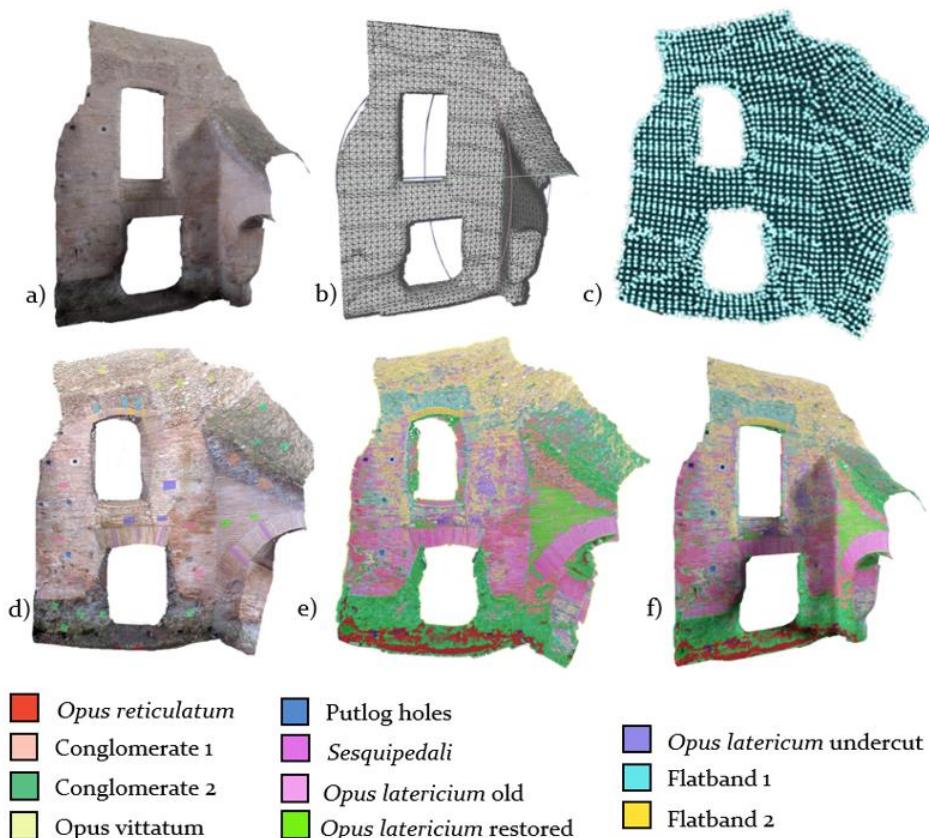


Figure 4.1. Schematic representation of the developed supervised classification methodology: 3D model of a portion of Circus Maximus Cavea in Rome, Italy, (a); 3D model after re-meshing (b); UV map (c); manually identified training areas on the unwrapped texture (d); supervised classification results (e); re-projection of the classification results onto the 3D model (f).

Part 2. Proposed methods

Considering the availability and reliability of segmentation methods applied to (2D) images and the efficiency of machine learning strategies, a new methodology was developed to assist cultural heritage experts in analysing digital 3D data. In particular, the approach presented hereafter relies on supervised and unsupervised machine learning methods for segmenting texture information of 3D digital models. Given a coloured 3D heritage, the proposed pipeline relies on the following steps:

1. generation of textured surface models, if not available yet;
2. creation of the orthoimages (for 2.5D geometries) or UV maps (for 3D geometries) (Figure 4.1 b-c);
3. classification of the orthoimages or UV maps following different approaches tailored to the case study (i.e., Random Forest or Clustering) (Figure 4.1 d-e);
4. re-projection of the 2D classification results onto the 3D heritage geometry (Figure 4.1 f).

Here follows a detailed explanation of the texture-based approach, while in Chapter 6 practical implementations are presented.

4.1. FROM 3D TO 2D

The proposed classification method works on the texture information of the 3D models. According to the geometry and complexity of the considered 3D object, it is possible to work with:

1. Orthophotos;
2. UV maps.

4.1.1 Orthophotos

Orthophotos are geometrically corrected photographs that has been 'orthorectified' such that the scale of the photograph is uniform. An ortho-photograph can be used to measure exact distances. For the classification of planar objects like walls, our approach suggests to export from the models the orthophoto or ortho-mosaic (a raster image made by merging orthophotos), classify it and finally re-maps the information onto the 3D geometry.

4.1.2 UV maps

UV maps are the flat representations of the surfaces of 3D models used to wrap textures easily. The process of creating a UV map is called UV unwrapping. The U and V refer to the horizontal and vertical axes of the 2D space, as X, Y, and Z are already being used in the 3D space (Figure 4.2).

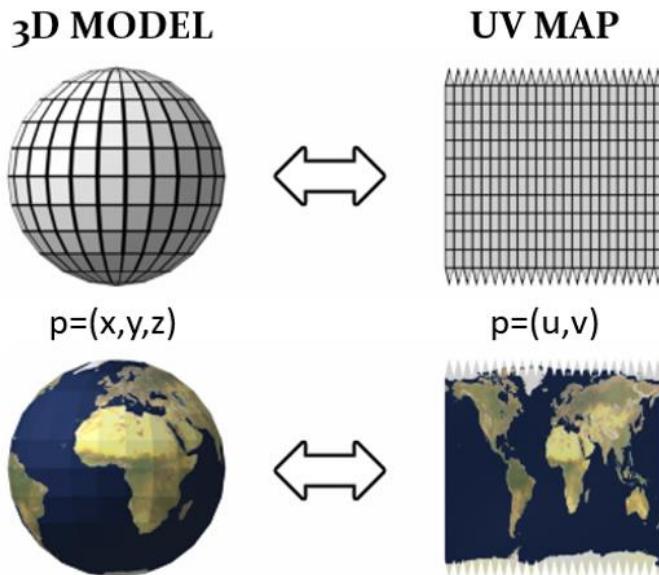


Figure 4.2. Schematic representation of a globe UV mapping.

The technological advances in computer graphics in the last decades have had a significant impact on fields such as animation, movies, video games, etc. leading to a broad diffusion of 3D applications software (e.g., 3ds Max, Maya, Cinema 4D, and Blender). In most 3D applications, UV maps can be generated automatically. Automatic mapping creates UVs for a mesh by attempting to find the best possible UV placement by projecting from multiple planes. These automatic tools are useful when you have complex shapes where the basic projections don't produce useful UVs (Section 6.3). Automatic mapping is an excellent place to start from, but manual editing of the UVs can facilitate the subsequent analysis with machine learning strategies.

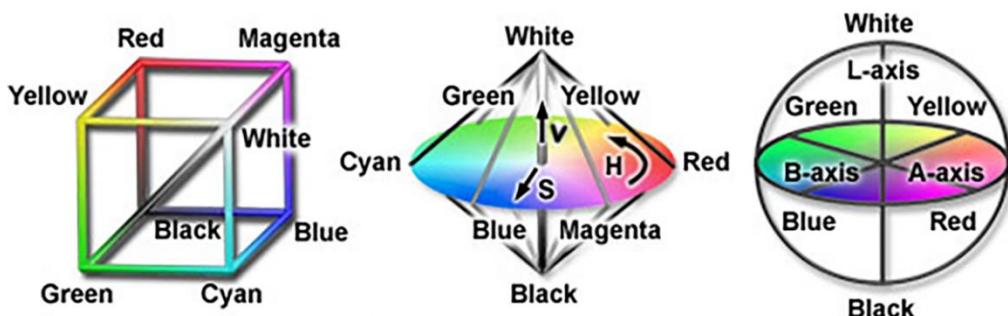
To simplify the unwrapping procedure, it is beneficial applying as a first step a remeshing to the 3D models. UV maps are then generated by adjusting and optimising seam lines and overlaps. Seams are the edges (or breaks) in geometry between UV islands and are inevitable for most models. Seams could cause problems during the classification process; areas of the models belonging to the same class may be split into different islands. These discontinuities between UV islands could bring to a misclassification. To avoid these kinds of problems the users can command the UV unwrapper to cut the mesh along edges chosen following the shape of the case study.

Once created, the UVs are textured using the original textured polygonal model (as vertex colour or external texture). This way, the radiometric quality is not compromised despite the remeshing phase.

The generation of textured UV maps allows us to classify in one single step complex objects, instead of creating various orthoimages from different points of view (Campanaro et al., 2016).

4.2 COLOUR SPACES

When we consider colour image segmentation, choosing a proper colour space becomes an important issue (Bora and Gupta, 2014). Different colour spaces present colour information in different ways that make some calculations more convenient and provide a way to identify colours that is more intuitive. Several colour representations are currently in use in colour image processing. The most common is the RGB, but also HSV and CIE L*a*b* are frequently chosen colour spaces (Sural et al. 2002; Jurio et al., 2010) (Figure 4.3). In the RGB colour space, for example, shadowed areas have most likely very different characteristics than areas without shadows. In the HSV colour space, the hue components of areas with and without shadow are more likely to be similar: the shadow primarily influence the value, or the saturation component, while the hue - indicating the primary "colour" without its brightness and diluted-ness by white/black - should not change so much. Another popular option is the LAB colour space. In the L*a*b* colour space, L* indicates lightness and a* and b* are chromaticity coordinates. The a* and b* coordinates are red/green and yellow/blue axis. Again, ignoring the L channel (Luminance) makes the algorithm more robust to lighting differences. According to the needs of the classification and case study characteristics in our experiments (Chapter 6) it was used or the colour component b* of the Colour space La*b* or the RGB values.



*Figure 4.10. Diagram of RGB, HSV and CIE L*a*b colour spaces (Russ et al., 2018).*

4.3. SUPERVISED LEARNING CLASSIFICATION

The 2D classification method relies on different machine learning models embedded in WeKa (Witten et al., 2016) coupled with the Fiji distribution of ImageJ, an image processing software that exploits WeKa as an engine for machine learning models (Schindelin et al., 2012). The method combines a collection of machine learning algorithms (random tree, random forest, etc.) with a set of selected image features to produce pixel-based segmentations. The features available can be categorised as:

Chapter 4. Texture-based classification approach

- *edge detectors*, which aim at indicating boundaries of objects in an image (e.g., Laplacian and Sobel filters, difference of Gaussians, Hessian matrix eigenvalues and Gabor filters);
- *texture filters*, to extract texture information (including filters such as minimum, maximum, median, variance, entropy, structure tensor, etc.);
- *noise reduction filters*, such as Gaussian blur, Bilateral filter, Anisotropic diffusion, Kuwahara and Lipschitz;

Besides, for coloured images, the hue, saturation, and brightness are also part of the features.

All the available classifiers in WeKa are based on a decision tree learning method. In this approach, during the training, a set of decision nodes over the values of the input features (e.g. “feature x is greater than 0.7?”) are built and connected in a tree structure. This structure represents a complex decision process over the input features. The result of this decision is a value for the label that classifies the input example. During the training phase, the algorithm learns these decision nodes and connects them. Among the different approaches, we achieved the best results in terms of accuracy exploiting the Random Forest method (Section 2.2.2).

For each case study, the RF was trained by giving in input the manually annotated orthophotos or UVs. Image annotations highlight and label a particular object by manually outlining. There are many types of annotation techniques such as the ones based on bounding boxes (Figure 4.4a), pixel-level labelling (Figure 4.4b), or polygonal selection (Figure 4.4c).

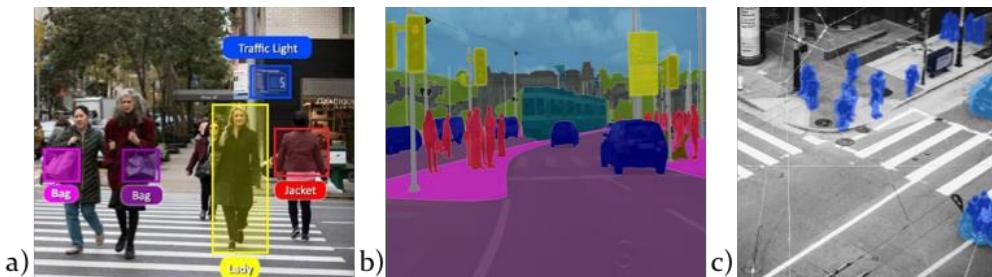


Figure 4.4. Image annotation techniques: bounding box (a); pixel-level labelling (b); polygonal annotation (c).

The choice of one method over another depends on the requirements and types of images. For all the case studies considered in this thesis (Chapter 6) the annotations were done following the polygonal selection techniques. In this way, just some significant and well-distributed portions of the textures were rapidly highlighted (Figure 4.1d), instead of labelling entire datasets like for deep learning approaches.

The first time the training process starts, the features of the input image are extracted and converted to a set of vectors of float values (WeKa input). This step can take some time depending on the image sizes, the feature number, and the computational power of the computers. After the training procedure, the

Part 2. Proposed methods

classifier gives as output a prediction about the classification of the entire orthophoto or UV map.

4.4. UNSUPERVISED LEARNING CLASSIFICATION

The unsupervised segmentation approach is performed using the k-means clustering plugin of ImageJ or Fiji (ImageJ K-means plugin, 2019). The algorithm (Section 2.2.3) performs pixel-based segmentation of multi-band images. Each pixel in the input image is assigned to one of the clusters. Values in the output image represent the cluster number to which the original pixel is assigned. Before starting the elaboration, the operator decides the number K of classes the image has to be divided into and the cluster centre tolerance.

4.5. SURFACES COMPUTATION

Among the various uses of the 3D classification results, there's the possibility to compute the areas that the different segmented classes occupy. This kind of data can be beneficial for quantity surveying, overcoming the performance of the surveyors concerning both the survey speed and the measurement accuracy.

Starting from the classified orthophotos or UV Maps, it is possible to estimate the quantity that each class occupies over the total area investigated. In particular, the percentage is calculated as a comparison between the number of pixels classified as Class-X and the total number of pixels in the classified image (subtracting the background). These percentages can then be transformed in surface measures, as we are working with scaled 3D models.

While the orthophotos are directly measurable, the measuring accuracy computed on the UV maps could be affected by errors due to distortions generated during the UV unwrapping. Therefore, it becomes essential working on equal-area projection UVs. The software Blender has an interesting tool called "Average island scale" that takes the selected faces and averages the scale of them, so they match according to the 3D view. For instance, if one face is scaled way down in the UV editor compared to its size in 3D space, it will get an appropriate scaling in relation to other selected parts of the mesh.

4.5.1 Quantity survey validation

A 3D model of the Rubik cube (Figure 4.4 a) has been taken into consideration to validate the quantity survey procedure described above.

The assessment can be summarized in the following step:

1. From the textured model, three different UV maps have been generated:
 - an automatic one (Figure 4.4 b);

Chapter 4. Texture-based classification approach

- an optimised version of the automatic one, coming from the application of the "average island scale" tool to get an equal-area projection (Figure 4.4 c);
 - a guided UV map created choosing a planar projection (Figure 4.4 d).
2. The three different UVs have been classified through a clustering procedure.

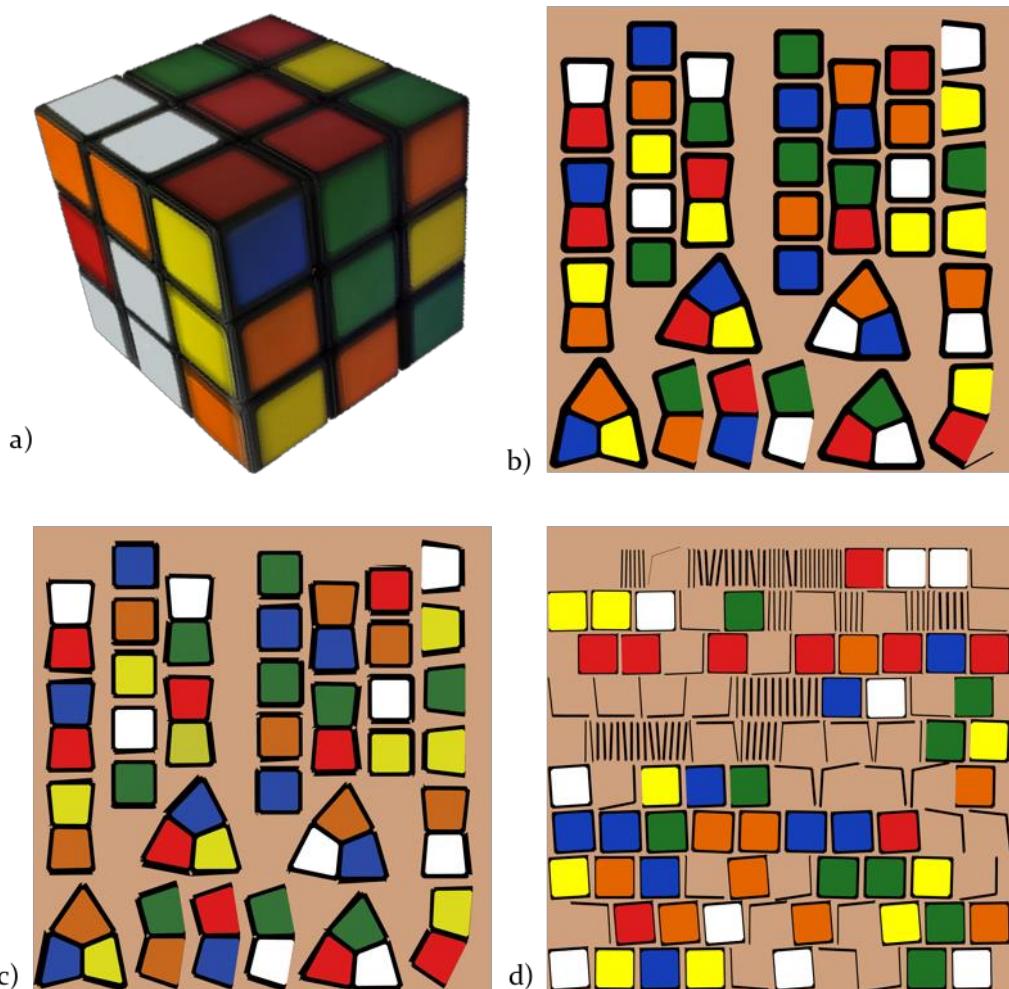


Figure 4.4. 3D model of the Rubik cube (a); automatic generated UV map (b); optimized UV map (c); planar projected UV map (d).

3. For all the UV maps, the percentage of each class (the six colours more the black trim) is calculated and multiplied for the total surface of the cube (known as a sum of the triangle areas of the mesh) (Table 4.1).
4. The measures coming from the segmented UV maps are compared with the actual surfaces of the cube (ground truth).

Concluding, as each coloured class of the cube occupies 23 cm^2 , it has been verified that to get accurate surface quantifications is beneficial working on equal-areas UV maps or planar projection (Table 4.1).

Part 2. Proposed methods

Ground truth: cube surface = 197cm ² /coloured areas surface = 23 cm ²						
CLASS	% OCCUPIED			cm ² OCCUPIED		
	AUTO. UV	OPTIM. UV	PLANAR UV	AUTO. UV	OPTIM. UV	PLANAR UV
black trim	37.66 %	32.18 %	27.42 %	74.18 cm ²	63.40 cm ²	54.02 cm ²
white	10.20 %	12.77 %	12.02 %	20.09 cm ²	25.15 cm ²	23.68 cm ²
red	10.40 %	11.26 %	12.36 %	20.49 cm ²	22.18 cm ²	24.35 cm ²
blue	10.01 %	10.95 %	12.00 %	19.73 cm ²	21.57 cm ²	23.64 cm ²
yellow	10.17 %	11.02 %	11.90 %	20.03 cm ²	21.71 cm ²	23.44 cm ²
orange	10.37 %	11.59 %	12.07 %	20.42 cm ²	22.83 cm ²	23.77 cm ²
green	11.20 %	11.90 %	12.23 %	22.05 cm ²	23.45 cm ²	24.10 cm ²
AVERAGE				20.47 cm²	22.82 cm²	23.83 cm²

Table 4.1. Comparison of the areas computed from the automatic clustering of the different UV maps. Optimez UV maps allow a more accurate quantification of areas.

REFERENCES

- Bora, D.J. and Gupta, A.K., 2014. A new approach towards clustering-based colour image segmentation. *International Journal of Computer Applications*, 107(12).
- Campanaro, D.M., Landeschi, G., Dell'Unto, N. and Touati, A.M.L., 2016. 3D GIS for cultural heritage restoration: A 'white box'workflow. *Journal of Cultural Heritage*, 18, pp.321-332.
- Imagej K-means plugin, 2019. [Online] Available at: <http://ij-plugins.sourceforge.net/plugins/segmentation/k-means.html> [Accessed January 27th 2020].
- Jurio, A., Pagola, M., Galar, M., Lopez-Molina, C. and Paternain, D., 2010. A comparison study of different color spaces in clustering-based image segmentation. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pp. 532-541. Springer, Berlin, Heidelberg.
- Russ, J.C., Parry-Hill, M. and Davidson M.W., 2018. *Molecular expression*. [Online] Available at: <https://micro.magnet.fsu.edu/primer/digitalimaging/russ/colorsplaces.html> [Accessed January 27th 2020].
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B. and Tinevez, J.Y., 2012. Fiji: an open-source platform for biological-image analysis. *Nature methods*, 9(7), p.676.
- Sural, S., Qian, G. and Pramanik, S., 2002, September. Segmentation and histogram generation using the HSV color space for image retrieval. In *Proceedings. International Conference on Image Processing*, Vol. 2, pp. 589-592. IEEE.
- Witten, I.H., Frank, E., Hall, M.A. and Pal, C.J., 2016. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

CHAPTER 5

Geometry-based classification approach

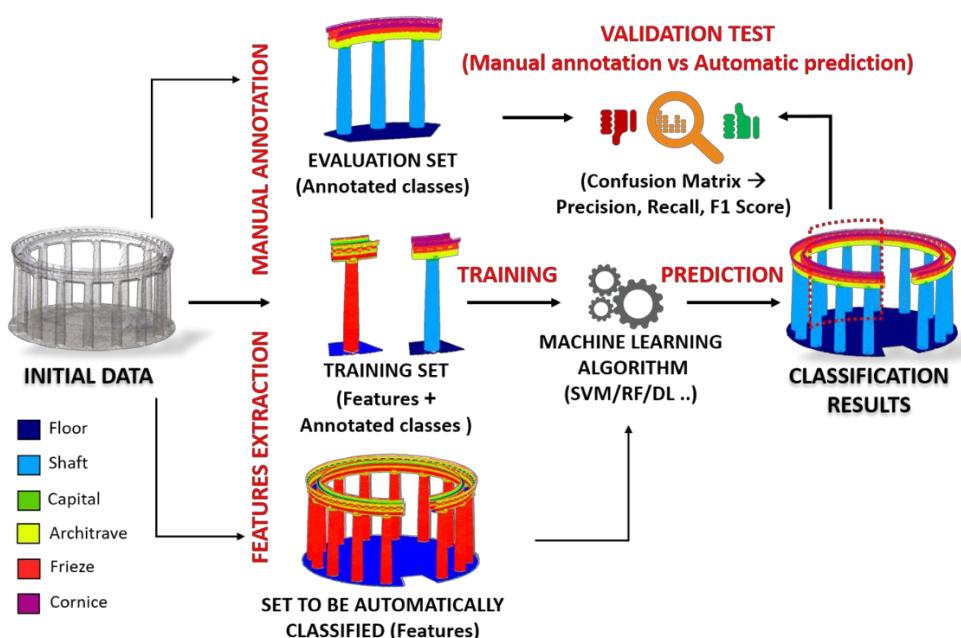


Figure 5.1. Supervised machine learning workflow. The laser scanning dataset of the Mausoleum of Cesare Battisti in Trento (Italy) is taken as an example.

Part 2. Proposed methods

For some case studies, it may happen that the colorimetric features are not available or enough to perform texture-based classification tasks (Chapter 4). To bridge this gap, a second classification approach based on the point cloud geometric characteristics has been developed.

The pipeline followed, similarly to the texture-based one (Chapter 4), is based on a supervised machine learning approach. The classification process (Figure 5.1) consists of five main steps:

- manual annotation (Section 5.1);
- feature extraction (Section 5.2);
- model training (Section 5.3);
- prediction (Section 5.3);
- validation (Section 5.4).

In the following table, the elaboration times for each step of the classification process are compared.



Table 5.1. Comparison of the elaboration times for the different classification phases.

5.1 MANUAL ANNOTATION

Annotation in machine learning is the process of manually labelling data, which could be in the form of text, images, audio, points etc. It is tedious and time-consuming work, but fundamental as to train machine learning classifiers and make them functional, as to give a semantic to the analysed data.

In our experiments, for each case study, small but significant samples of the entire datasets have been selected to be manually classified. To make the classification possible, we must make sure that the training data are representative of the entire scenarios, hence the samples must contain all the classes under investigation over the whole datasets. The manual annotation is performed using the segment tool within the open-source software Cloud Compare. After the manual segmentation, a class index is associated with every segment of the point clouds.

5.2 FEATURE EXTRACTION

What are the features and why we need to extract them? As already mentioned in Chapter 2.2, all machine learning algorithms use some input data to learn and predict outputs. This input data, in addition to the manual annotation, comprise

Chapter 5. Geometry-based classification approach

features, which are information extracted from the input data to simplify the learning between the input and output data. Algorithms require features with some specific characteristic to work correctly. When we work on point cloud classification, features are generally geometric or radiometric attributes that are useful to highlight the heterogeneity between the classes. Based on the training data, the machine learning classifier gives as output a semantic segmentation prediction for the entire datasets. So, the choice of the features directly influences the predictive model and the results you can achieve. Without relevant features, you can't train an accurate model/classifier, no matter how sophisticated the machine learning algorithm is.

For the classification experiments, different sets of features have been used, depending on the case study and the chosen approach (Machine / Deep Learning). It has been combined the use of:

- Decentralised coordinates: used to represent the local geometry around a point as a patch of k-number of nearest points. To decentralise the coordinates, the minimum x, y, z values are subtracted within each sequence, and the sequences are sorted with respect to the decentralised coordinate values. Decentralised coordinates have been used just for the deep learning approaches (Grilli et al., 2019b).
- Radiometric values: for some case studies, the colorimetric values associated with the points of the cloud were beneficial for the understanding of the datasets.
- Geometric features: including (i) covariance features (Section 5.2.1), (ii) normal based features and (iii) height-based features (Table 5.2).

5.2.1 The covariance features

The covariance features (Chehata et al., 2009) are widely used in segmentation and classification procedures because of their capability to provide deep knowledge on the geometrical structure of the reconstructed scene (Weinmann et al. 2013, Hackel et al., 2016, Weinmann et al. 2017). Despite their widespread use in the geospatial field, there is no literature about their application to architectural case studies.

These features are shape descriptors obtained as a combination of eigenvalues ($\lambda_1 > \lambda_2 > \lambda_3$) extracted from the covariance matrix (Blomley et al., 2014). The covariance matrix can be considered as 3D tensors containing geometrical information about the point distribution within a neighbourhood. Using a statistical analysis, the Principal Component Analysis (PCA), it is possible to extract from the covariance matrix the three eigenvalues $\lambda_1, \lambda_2, \lambda_3$ representing the local 3D structure and measuring the variation of the local point set along the direction of the corresponding eigenvector (Figure 5.2).

Part 2. Proposed methods

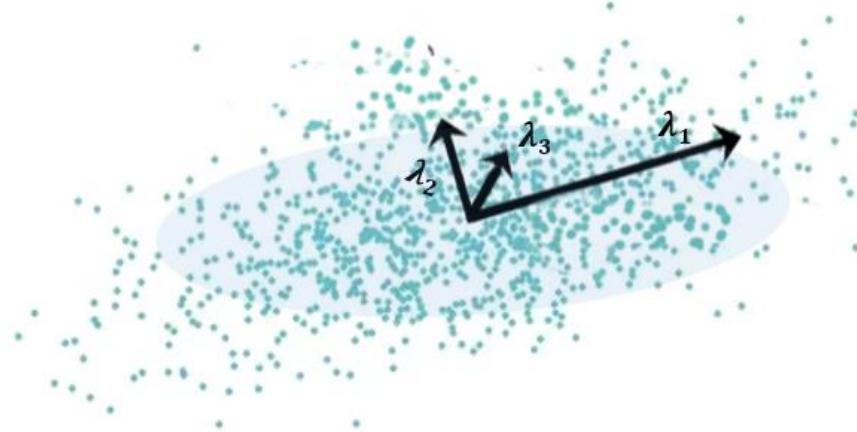


Figure 5.2. Local distribution of the points along the three main directions.

The combination of these features values highlights the main linear (1D), planar (2D) or volumetric (3D) structure of the point cloud in the neighbourhood. Their definition is presented in Table 5.2.

TYPE	NAME	FORMULA
COVARIANCE FEATURES	Linearity	$L_\lambda = \frac{\lambda_1 - \lambda_2}{\lambda_1}$ (1)
	Planarity	$P_\lambda = \frac{\lambda_2 - \lambda_3}{\lambda_1}$ (2)
	Sphericity	$S_\lambda = \frac{\lambda_3}{\lambda_1}$ (3)
	Omnivariance	$O_\lambda = \sqrt[3]{\prod_{j=1}^3 \lambda_j}$ (4)
	Anisotropy	$A_\lambda = \frac{\lambda_1 - \lambda_3}{\lambda_1}$ (5)
	Eigenentropy	$E_\lambda = -\sum_{j=1}^3 \lambda_j \ln (\lambda_j)$ (6)
	Sum of Eigenvalues	$\Sigma_\lambda = \sum_{j=1}^3 \lambda_j$ (7)
	Surface Variation	$C_\lambda = \frac{\lambda_3}{\Sigma \lambda}$ (8)
NORMAL-BASED FEATURE	Verticality	$V = 1 - n_z$ (9)
HEIGHT-BASED FEATURE	Height	Z Coordinate

Table 5.2. Considered geometric features.

Different strategies may be applied to recover the local neighbourhood for points belonging to a 3D point cloud. It can generally be computed as a sphere or a cylinder with a fixed radius or be described by the number of the $k \in N$ nearest neighbours (Weinmann et al., 2015).

Chapter 5. Geometry-based classification approach

In the presented research the features have been calculated on spherical neighbourhoods at various radius sizes (Brodu and Lague, 2012), to explore different responses in function to the different geometric properties of the heritage monuments (Further details in Section 5.5). Cloud Compare software was used for feature extraction (Hackel et al., 2016).

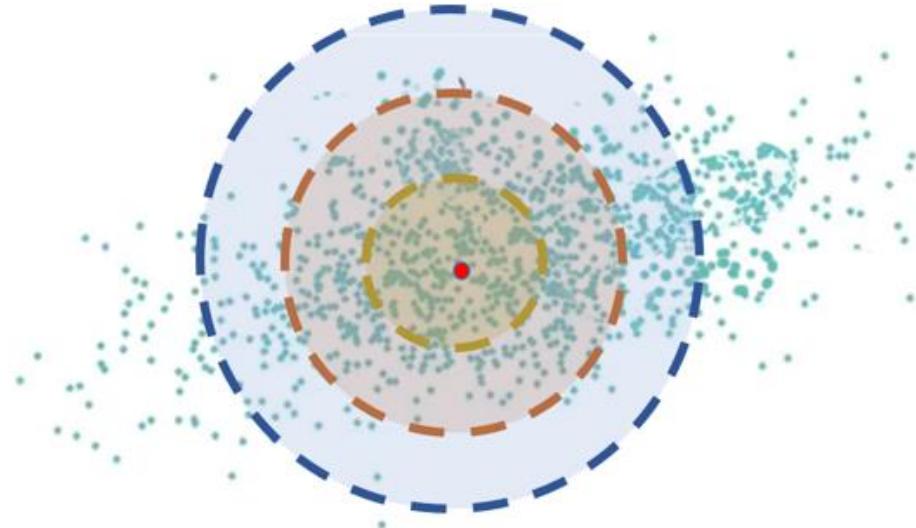


Figure 5.3. PCA analysis of the main distribution of the points using spheres with different radii.

5.3 MODEL TRAINING

After preparing the training set with annotated classes and extracted features, different predictive models (machine and deep learning classifiers) were trained, to extend the semantic segmentation from small portions to entire datasets (prediction phase).

5.3.1 Random Forest classifier

To test the classification capabilities of the Random Forest algorithm (already mentioned in Section 2.2.2), a RF implementation available in the Scikit-learn Python library (version 0.21.1) (Pedregosa et al., 2011) was used. Two parameters needed to be set to produce the forest trees: the number of decision trees to be generated (N_{tree}) and the number of variables to be selected and tested for the best split when growing the trees (M_{try}) (Belgiu et al., 2016). During the training process, the N_{tree} and M_{try} were tuned considering the best F1-score (accuracy metric described in Section 5.4) computed on the evaluation set.

5.3.2 OvO classifier

Classification using the One-versus-One (OvO) classifier from dlib C++ library was also carried out (King, 2009). OvO converts a group of binary classifiers into a multiclass classifier. It works training the binary classifiers in a one vs. one trend. In the case of N possible classes, it trains $N*(N-1)/2$ binary classifiers, which are then employed for the identification of the classes on the test sample.

Part 2. Proposed methods

5.3.3 1D AND 2D CNN

Two methods belonging to the category of Convolutional Neural Networks (CNN) (Fukushima et al., 1980) are also proposed. As written in Section 2.2.4, CNNs are networks specialised in processing data that have a grid-like topology, such as images. The layers of a CNN consist of an input layer, an output layer and a hidden layer that includes multiple convolutional layers, pooling layers, fully connected layers and normalization layers. The CNNs we propose are:

- 1D CNN: consists of 1 input layer, 2 convolutional layers, 3 dense layers, 1 maximum pooling layer, 1 global average pooling layer, and 1 dropout layer.
- 2D CNN: composed with 1 input layer, 4 2D convolutional layers, 2 2D max pooling layer, 3 dropout layers, 1 flatten layer, and 2 dense layers.

5.3.4 Bi-LSTM

Finally, a deep learning approach, based on a Recurrent Neural Network (RNN) was run (Rumelhart et al., 1988). RNNs are commonly used for modelling sequential data. A data is sequential if the building blocks in a dataset are not independent from each other. The most common applications for RNNs are handwriting or speech recognition and translation. The RNN used consists of five layers: sequence input layer, Bidirectional Long Short-Term Memory layer (Bi-LSTM) with 200 hidden units, fully connected layer, softmax layer, and classification layer. Each point of the clouds was described with a sequence generated with its surrounding points. These sequences are expected to represent the geometry around each point in a better way when compared to a single feature vector representation.

5.4 VALIDATION

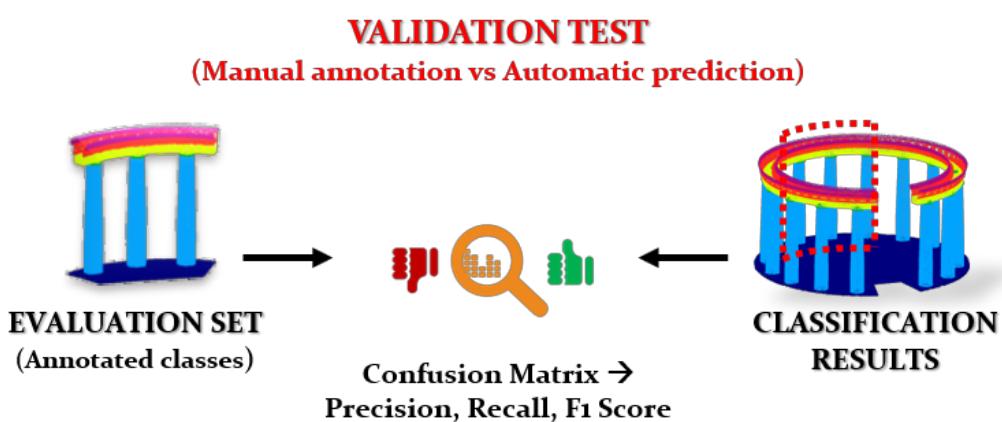


Figure 5.4. Validation test workflow. Mausoleum of Cesare Battisti in Trento (Italy).

For all case studies a small portion of the point clouds was taken into consideration as evaluation dataset, in order to assess the classification results. On this dataset, the label predicted by the classifier are compared with the same

Chapter 5. Geometry-based classification approach

previously manually annotated (Figure 5.4). The number of correct and incorrect predictions are summarized with count values and broken down by each class inside a confusion matrix, a specific table layout that allows the visualization of the performance of the algorithm (Table 5.3). Each row of the matrix represents the instances in an actual class (ground truth), while each column represents the instances in a predicted class. The Confusion matrix is not a performance measure as such, but all the accuracy metrics are based on the numbers inside it:

- Precision: it is a ratio of the total detection by the classifier.

$$Precision = \frac{Tp}{Tp + Fp} \quad (10)$$

- Recall: it is a ratio of the correct detection over the total number of test samples.

$$Recall = \frac{Tp}{Tp + Fn} \quad (11)$$

- F1 score: it is used to compare the performance of the predictive model. It considers both the precision and recall values to compute the measures.

$$F1\ score = 2 * \frac{Recall * Precision}{Recall + Precision} \quad (12)$$

Where Tp = True positive, Fp = False positive, Fn = False negative.

PREDICTED CLASSES										
CLASS	Floor	Shaft	Capital	Architr.	Frieze	Cornice	Prec.	Recall	F1	
Floor	33949	261	32	12	0	15	99%	100%	100%	
Shaft	13	46075	0	0	0	0	100%	99%	99%	
Capital	0	221	1318	126	0	0	79%	94%	86%	
Architr.	0	0	48	9802	166	0	98%	97%	98%	
Frieze	0	0	0	117	8652	46	98%	92%	95%	
Cornice	0	0	0	0	596	4795	89%	99%	94%	
AVERAGE							95%	95%	95%	

Table 5.3. Confusion matrix relative to the Mausoleum of Cesare Battisti in Trento (Italy).

As an example, we can consider the above Confusion Matrix, for the class “capital”:

- True positive: diagonal position, (capital, capital).
- False positive: sum of column “capital” (without main diagonal).

Part 2. Proposed methods

- False negative: sum of row “capital” (without main diagonal).

In order to find one value for the entire model, Precision, Recall and F1 score are first computed for each class using the above formula, then the average over all the classes is considered.

In short, Recall gives information about a classifier’s performance with respect to false negatives (how many did we miss), while Precision gives information about its performance with respect to false positives (how many did we catch). If the aim is minimising False Negatives, we would want our Recall to be as close as possible to 100% without negatively impacting precision, while if we prefer minimising False Positives, then our focus should be to make Precision as close as possible to 100%.

5.5 FEATURE SELECTION

As said in Section 5.2, a critical part of the success of a Machine Learning project consists of a good selection of the features used for the training.

It has been verified that the covariance features (Section 5.2.1) can behave in completely different ways when computed at different search radii (Figure 5.5). Consequently, the so-called multi-scale approach, which combines features extracted at different neighbourhood sizes, compared with the single-scale approach, was proved to be beneficial for the classification results (Brodu and Lague, 2012; Hu et al., 2013; Niemeyer et al., 2014; Schmidt et al., 2014).

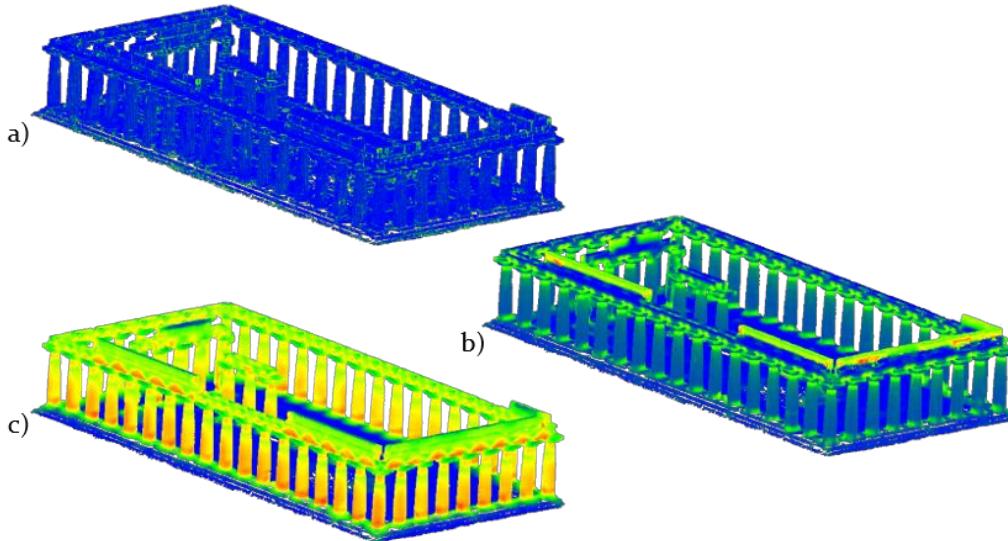


Figure 5.5. Different behaviors of the same covariance feature computed at different radius sizes: surface variation at 0.2m (a), 0.7m (b), 1.4m (c).

The main problem of using multi-scale neighbourhoods is their computational time, that grows as the density of the point clouds, the number of features to be extracted, and the size of the search radii increase. After this consideration, the research aims to:

Chapter 5. Geometry-based classification approach

- reduce the computation time necessary for extracting the features (Grilli et al., 2019a).
- identify a subset of features that performs well with different heritage datasets;
- improve the classification performances compared to a standard multi-scale approach.

Our feature selection framework can be summarized in Figure 5.6.

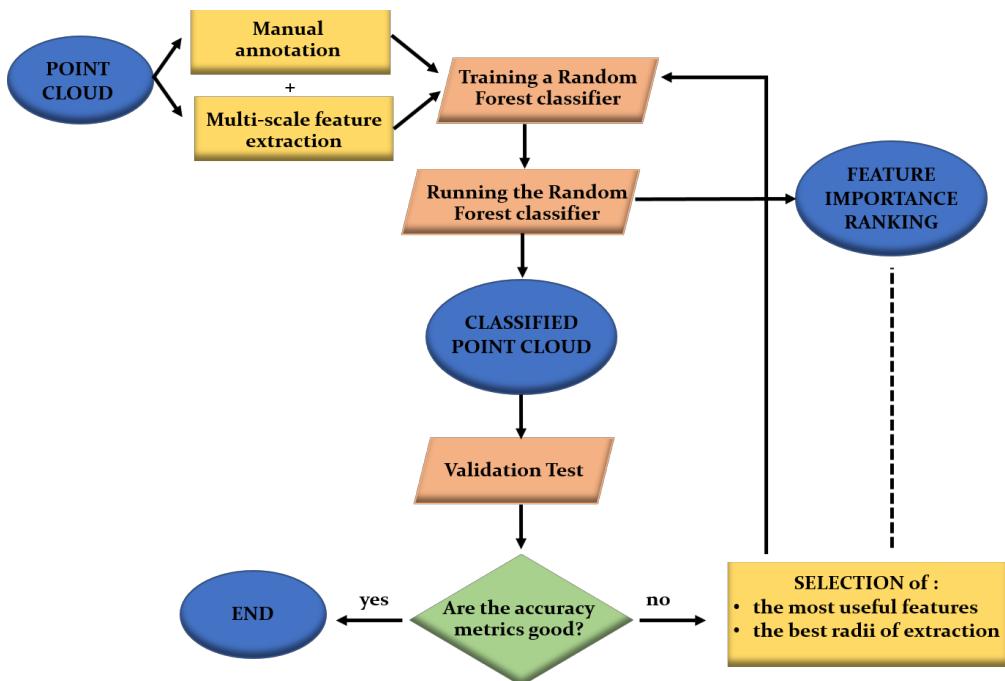


Figure 5.6. Feature selection workflow.

At first, the covariance features are extracted at increasing radii sizes (multi-scale approach). Secondly, a Random Forest classifier is trained and run to predict the classification. The reason why the RF classifier was chosen for this analysis is that RF offers a useful feature selection indicator. Specifically, it shows the relative importance or contribution of each feature in the prediction: it automatically computes the relevance score of each feature in the training phase, then it scales the relevance down so that the sum of all scores is 1. The feature importance ranking obtained tells us what variables are the most discerning between classes. Moreover, it is used for feature selection by iteratively removing low importance features. Finally, the results coming from different combinations of features are compared, relying on the confusion matrix scores previously mentioned.

Practical experiments on feature selection are shown in Section 7.1.

5.6 Computational Geometry Algorithms Library (CGAL)

The development of our geometry-based approach was inspired by the supervised classification method implemented in The Computational Geometry Algorithms Library (CGAL) by Giraudot and Lafarge (2019).

CGAL works in combination with a RF classifier delivered by ETH Zurich (ETH Zurich Random Forest Template, 2015). Moreover, it provides some predefined features that can be extracted with a multi-scale approach. The number of scales that can be used is included within a range between one scale and 10 scales. When using one scale, the features are calculated at the smaller neighbourhood size possible, that is equal to the density of the points. With two scales, features are also extracted at double of the previous neighbourhood, and so on till 10 scales. Here follows the list of the available features:

- Distance to plane: it measures how far away a point is from a locally estimated plane;
- Eigenvalues: it measures one of the three local eigenvalues;
- Elevation: computes the local distance to an estimation of the ground;
- Height above: it computes the distance between the local highest point and the point;
- Height below: it computes the distance between the point and the local lowest point;
- Vertical dispersion: it computes how noisy the point set is on a local Z-cylinder;
- Vertical range: it computes the distance between the local highest and lowest points;
- Verticality: it compares the local normal vector to the vertical vector.

In addition, if available in the point cloud in input, the colour channels and the echo scatter (number of returns provided by most LIDAR scanners) are used among the features.

Two different case studies treated with CGAL will be presented in Chapter 7. Although quite satisfying results have been achieved for the first experiment (Section 7.1), the second dataset (Section 7.5) revealed different problems, due to his complexity. Considering not fully exhaustive the results achieved using CGAL, which is a powerful tool but designed for a geospatial environment, during the PhD it was developed an approach more specific for the architectural field.

Chapter 5. Geometry-based classification approach

REFERENCES

- Belgiu, M. and Drăguț, L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, pp.24-31.
- Blomley, R., Weinmann, M., Leitloff, J. and Jutzi, B., 2014. Shape distribution features for point cloud analysis-a geometric histogram approach on multiple scales. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2(3), p.9.
- Brodu, N. and Lague, D., 2012. 3D terrestrial lidar data classification of complex natural scenes using a multi-scale dimensionality criterion: Applications in geomorphology. *ISPRS Journal of Photogrammetry and Remote Sensing*, 68, pp.121-134.
- Chehata, N., Guo, L. and Mallet, C., 2009. Airborne lidar feature selection for urban classification using random forests. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 38 (Part 3).
- ETH Zurich Random Forest Template, 2015. [Online] Available at: www.prj.igp.ethz.ch/research/Source_code_and_datasets.html [Accessed May 15th 2019].
- Fukushima, K., 1980. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4), pp.193-202.
- Giraudot, S. and Lafarge, F., 2019. Classification. In *CGAL User and Reference Manual*. CGAL Editorial Board, 4.14 edition.
- Grilli, E., M. Farella, E., Torresani, A. and Remondino, F., 2019a. Geometric features analysis for the classification of cultural heritage point clouds. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XLII-2/W15, pp 541-548.
- Grilli, E., Özdemir, E. and Remondino, F., 2019b. Application of machine and deep learning strategies for the classification of heritage point clouds. In proceeding of *The international ISPRS Geospatial Conference 2019*, University of Tehran. (In press)
- 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. 3(3), pp. 177-184.
- Hu, H., Munoz, D., Bagnell, J.A. and Hebert, M., 2013, May. Efficient 3-d scene analysis from streaming data. In *2013 IEEE International Conference on Robotics and Automation*, pp. 2297-2304. IEEE.
- King, D.E. 2009. Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research* 10, pp. 1755-1758.
- 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.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J., 2011. Scikit-learn: Machine learning in Python. *Journal of machine learning research*, 12(Oct), pp.2825-2830.
- Rumelhart, D.E., Hinton, G.E. and Williams, R.J., 1988. Learning representations by back-propagating errors. *Cognitive modeling*, 5(3), p.1.
- Schmidt, A., Niemeyer, J., Rottensteiner, F. and Soergel, U., 2014. Contextual classification of full waveform lidar data in the Wadden Sea. *IEEE Geoscience and Remote Sensing Letters*, 11(9), pp.1614-1618.

Part 2. Proposed methods

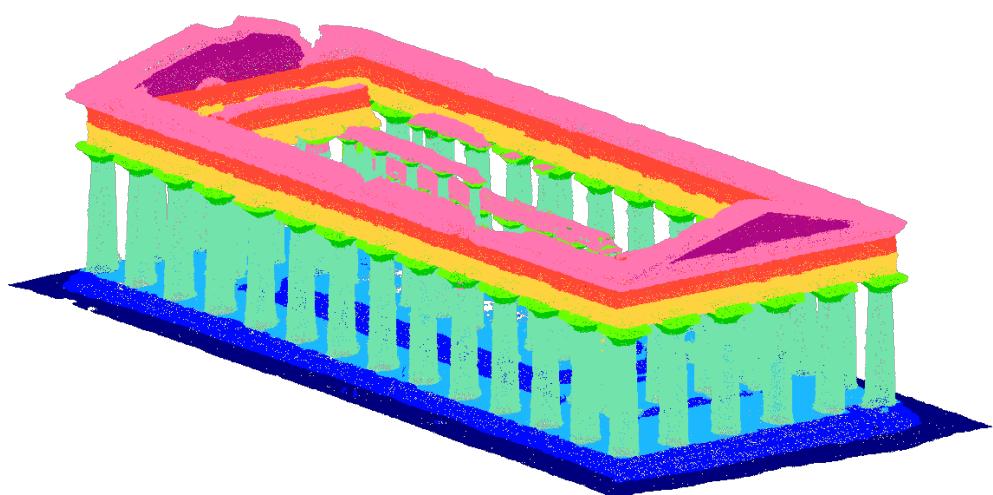
Weinmann, M., Jutzi, B. and Mallet, C., 2013. Feature relevance assessment for the semantic interpretation of 3D point cloud data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, II(5/W2), pp. 313-318.

Weinmann, M., Jutzi, B., Hinz, S. and Mallet, C., 2015. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing*, 105, pp.286-304.

Weinmann, M., Jutzi, B. and Mallet, C., 2017. Geometric features and their relevance for 3D point cloud classification. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4, p.157.

PART 3

Case studies



CHAPTER 6

Textured-based classification results

The following archaeological/architectural scenarios have been considered for testing the reliability of the texture-based approach presented in Chapter 4:

- The Pecile wall in Villa Adriana, Tivoli: an orthomosaic of a part of the Roman wall (60 m L x 9 m H) was classified and then re-projected onto the 3D digital model. The classification aimed to identify the different categories of *opus* (classical building techniques) it is made of, distinguishing original and restored parts (Section 6.1).
- The Sarcophagus of the Spouses exposed in Villa Giulia Museum in Rome: a late 6th century BC Etruscan sarcophagus (1,9 m L x 1,4 m H x 0,8 m D). It was chosen to test the reliability of the texture-based approach applied to an object with a complex topology. The classification aimed at identifying the surface anomalies, and quantifying the mimetic cement used to assemble the sarcophagus, once fragmented in many pieces (Section 6.2).
- The Bartoccini's tomb in Tarquinia: a 4th-century tomb composed of four rooms - a central one (ca. 5 m L x 3 m H x 4 m D) and three later ones (ca. 3 m L x 3 m H x 3 m D). The texture-based classification, held onto the available panoramic pictures, aimed at quantifying the deteriorated parts of the walls (Section 6.3).
- A small portion of porticoes in Bologna (ca. 8 m L x 13 m H x 5 m D): the classification, held at first on the UV map of the model, meant to identify its principal parts and architectural elements (Section 6.4).

6.1 PECILE WALL IN VILLA ADRIANA, TIVOLI

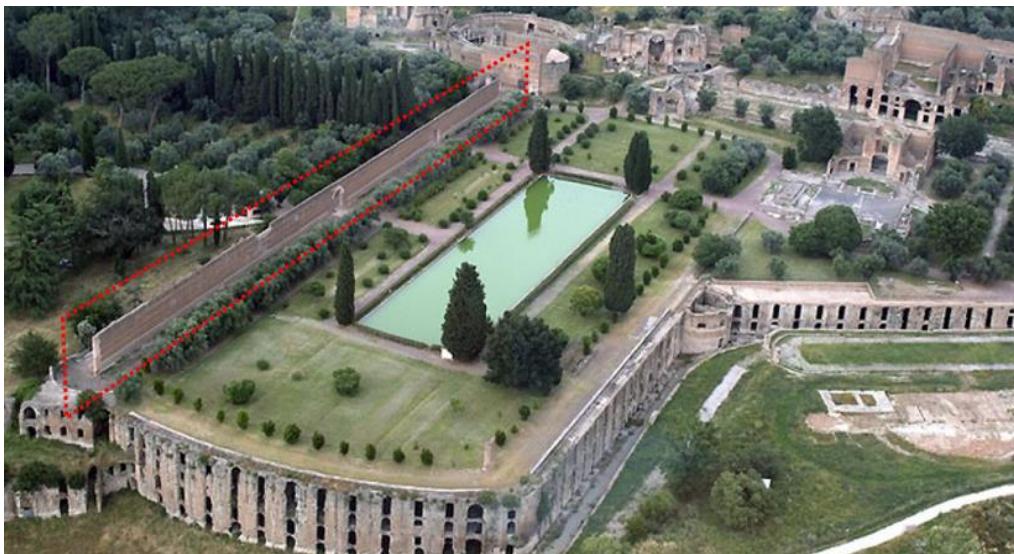


Figure 6.1. Aerial view of a part of Villa Adriana with the Pecile wall highlighted.

Hadrian's Villa (Villa Adriana) in Tivoli was built by Roman Emperor Hadrian at the foot of the Tiburtine Hills between 118 and 138 AD. The vast residential complex is extended over an area of about 120 hectares. The site includes buildings, baths, temples, barracks, theatres, gardens, fountains and *nymphaeums*, constructed in travertine, brick, lime, pozzolana and tufa. The Hadrian's Pecile was a monumental four-sided portico, delimiting a garden with a large swimming pool centrally located. A wall about 9 meters high and 100 meters long, once designed for walking in the sun and shade depending on the season, is what remains of the original quadriportico (Figure 6.1).

A digital model of the wall is available, as result of a photogrammetric survey conducted in summer 2016 in conjunction with the International Seminar of Museography at Villa Adriana - Piranesi Prix de Rome (Figure 6.2). This photogrammetric model was used for some classification experiments, designed to identify the Roman building techniques of the wall (Grilli et al., 2018).



Figure 6.2. A close view of the Pecile wall in Villa Adriana, Tivoli - Italy.

Chapter 6. Texture-based classification results

Different analyses have been conducted to assess the texture-based approach. At first, only a portion of the Pecile wall (4 m length × 9 m height) was considered (Figure 6.3a). From this portion of the model, three orthophotos exported at different scale were generated, manually annotated (Figure 6.3b) and automatically classified. With a 1:10 scale, the orthophoto was over-segmented (Figure 6.3c). On a 1:50 scale, some macro-areas were identified, but many details were lost (Figure 6.3e). The 1:20 scale (generally used for restoration purposes) turned out to be the optimal choice for the classification purposes. The classification was able to capture different details but at the same time to avoid the identification of the joints between the bricks (Figure 6.3d).

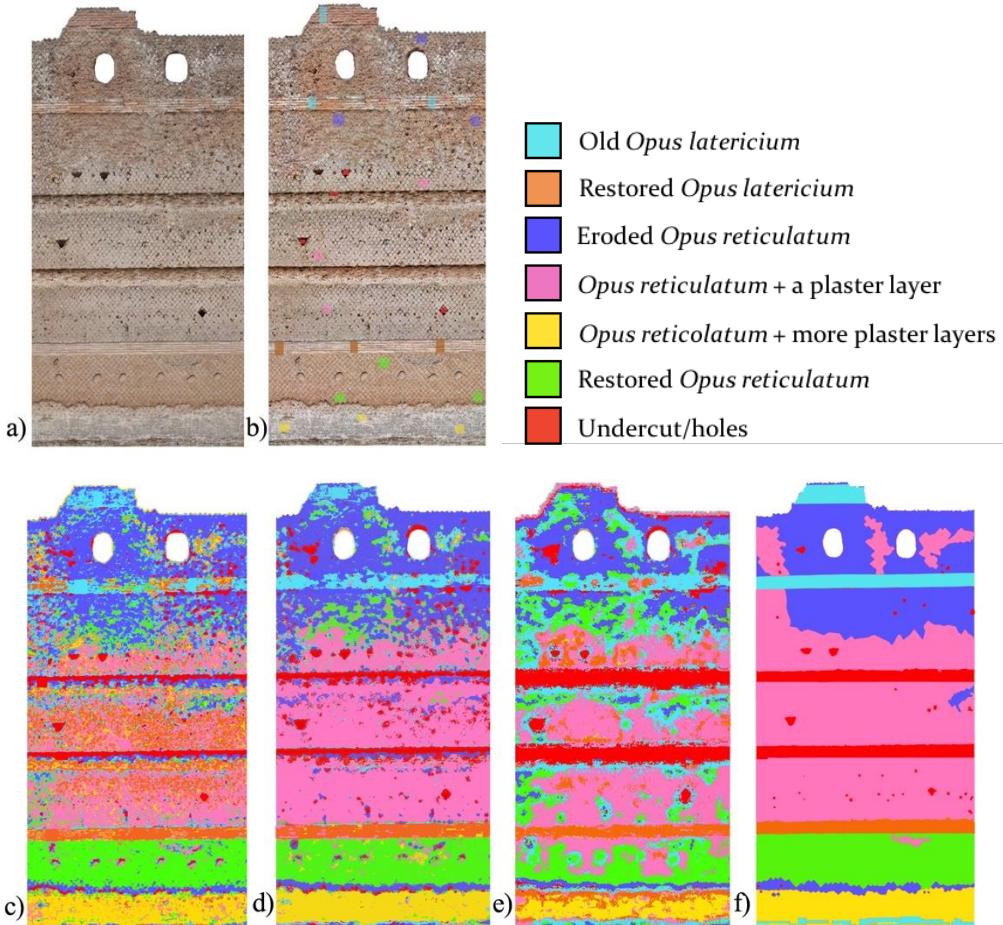


Figure 6.3. Orthophoto of a portion of the Pecile wall (4 m L×9 m H) (a); corresponding training samples (b); classification results obtained at different scales: scale 1:10 (c); scale 1:20 (d); scale 1:50 (e); manually segmented portion / ground truth (f).

Different predictive models available in Weka (Witten et al., 2016) were explored and compared starting from the same manually annotated orthoimage exported at 1:20 scale. All the different classification results were compared with the correspondent ground truth (the same portion manually segmented), and for each algorithm the overall accuracy was calculated (Table 6.1) as follow:

$$\text{Overall accuracy} = \frac{\text{N of pixel correctly classified}}{\text{total number of pixel}} \quad (13)$$

Part 3. Case studies

Out of all the tests performed with the different algorithms, the best overall accuracy obtained was around 70% using a Random Forest classifier.

Classifier	Overall accuracy	Time for processing
j48	44 %	22 s
Random Tree	46 %	15 s
RepTREE	47 %	33 s
LogitBoost	52 %	20 s
Random Forest	70 %	120 s

Table 6.1 Accuracy results and elapsed time for various classifier applied to an orthoimage at 1:20 scale.

To better identify the classification errors, a normalised confusion matrix was used (Table 6.2). From the table analysis, it was possible to understand that most failures in classification were in those classes where an overlap of plaster was present on the surface of the opus. However, it is believed that an expert should not consider as absolute the accuracy percentage without prior verification. In fact, comparing the classification handled by the operator and the one by the algorithm, it was found that the supervised method could identify more details and differences in the material's composition.

CLASS	Undercut	Restored Opus Lat.	Opus ret + plasters	Old Opus Latericum	Opus ret. + plaster	Restored Opus Ret.	Eroded Opus ret.	Prec.	Rec.	F1
Undercut	0.64	0.01	0	0.07	0.08	0.02	0.13	67%	55%	60%
Restored Opus Lat.	0.2	0.73	0.02	0.02	0.18	0.01	0	63%	85%	72%
Opus ret. + plasters	0.02	0.03	0.75	0.04	0.12	0.02	0.03	74%	91%	82%
Old Opus Latericum	0.08	0.02	0.01	0.43	0.2	0.05	0.21	43%	56%	49%
Opus ret. + plaster	0.06	0.05	0.01	0.08	0.66	0.05	0.08	67%	47%	55%
Restored Opus Ret.	0.02	0.01	0.01	0.01	0.07	0.83	0.06	82%	78%	80%
Eroded Opus ret.	0.15	0.01	0.02	0.12	0.08	0.09	0.52	53%	50%	51%
AVERAGE								64%	66%	64%

Table 6.2. Normalised Confusion Matrix and accuracy metrics for the Pecile wall classification.

Finally, using the same model training, the classification was extended to a more significant part of the wall (60 m L × 9 m H) (Figure 6.4a). To classify 540 m² of surface (Figure 6.4b) the process took about one hour. Considering that the operator needed four hours to manually classify a smaller part (24 m²), with the supervised approach, it was possible to get more accurate results in a shorter

Chapter 6. Texture-based classification results

time. In addition, instead of using random colours, it is possible to represent the classification results with a dedicated legend, commonly requested for degradation analysis tables (Figure 6.4d).

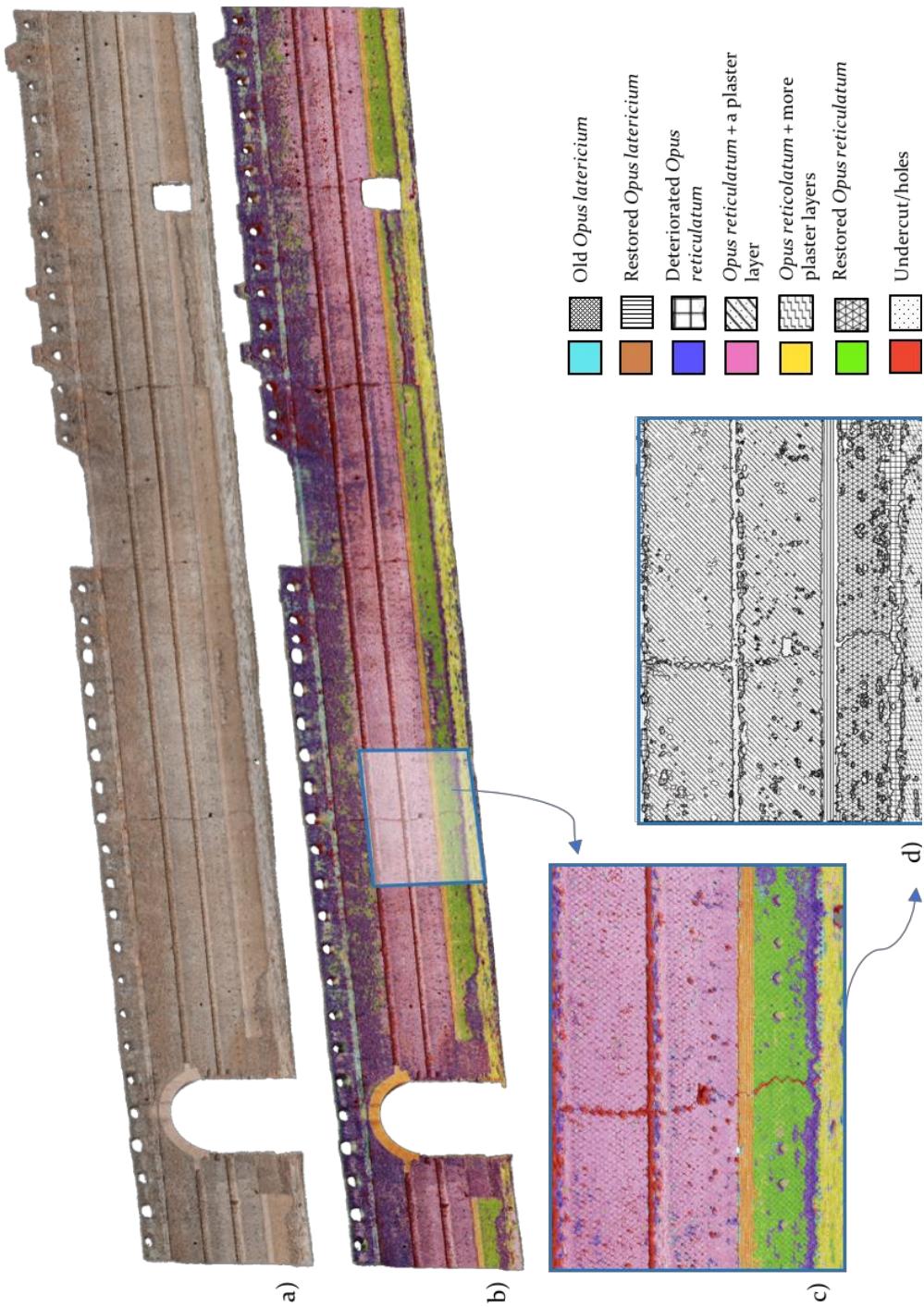


Figure 6.4. The original (a) and classified (b) model of the Pecile wall long ca 60 m. A closer view is also reported to better show the classification results with random colours (c) or dedicated symbols (d).

6.2 SARCOPHAGUS OF THE SPOUSES

The Sarcophagus of the Spouses is a late sixth-century BC Etruscan anthropoid (human-shaped). Made of painted terracotta, it depicts a married couple reclining at a banquet together. The sarcophagus, which would have originally contained cremated human remains, was discovered in 1881 during archaeological excavations in the Banditaccia necropolis of ancient Caere (now Cerveteri, Italy). The Etruscan masterpiece was found broken into more than 400 pieces. It was then reassembled and joined using a mimetic cement to fill the gaps among the different pieces. It is now exposed in the National Etruscan Museum of Villa Giulia, Rome (Figure 6.5).



Figure 6.5. The Sarcophagus of the Spouses, Villa Giulia museum – Rome.

In 2013, digital acquisitions and 3D modelling of the sarcophagus, based on different technologies (photogrammetry, TOF, and triangulation-based laser scanning) were conducted by project partners (FBK-3DOM, Cineca, CNR-ISTI, CNR-ITABC) to deliver a highly detailed photo-realistic 3D representation of the Etruscan masterpiece for successive multimedia purposes (Figure 6.6a) (Menna

Chapter 6. Texture-based classification results

et al., 2016). The segmentation task aimed to detect the decays of the pottery surface and to test the reliability of the method on a heritage objects with more complex topology and few chromatic differentiations on the texture.

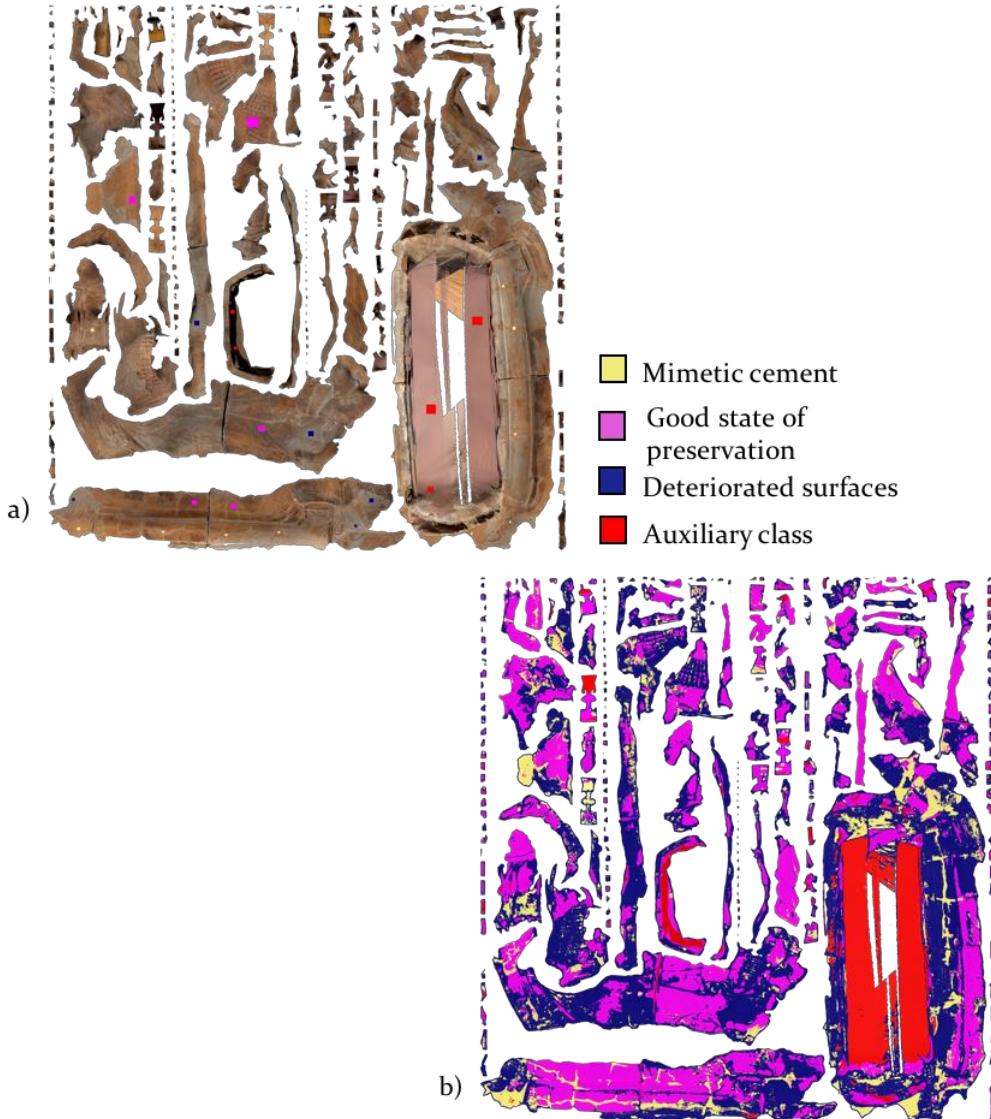


Figure 6.6. Manually annotated training areas on the unwrapped texture of the sarcophagus (a); related classification result (b).

The manual identification and annotation of the training patches took around 15 minutes and was accomplished with the support of restoration experts (Figure 6.6b). For the classification scope, three main categories and two ancillary ones were identified. After the manual annotation, about one hour of processing was necessary to extract the features, train the RF, and predict the classification on the entire texture (Figure 6.6c). Once the UV map was classified was then mapped onto the available 3D geometry (Figures 6.7). The segmented 3D model highlighted every single detail of the masterpiece assembly; fractures were distinguished from engraving, and the different grades of conservation were also identified. The sustaining legs of the sarcophagus were excluded from the classification, as they are the only parts where pigment decorations are clearly visible; thus, their analysis was outside the segmentation scope. From the

Part 3. Case studies

classification output, we could also calculate the percentage that each label occupies; from the results, we have that 12% of the entire surface of the object (i.e., 3D model) is composed of mimetic cement. As the overall surface of the 3D model is 6.8 m^2 , it means that approximately 0.8 m^2 are reconstructed parts (Figure 6.8).



Figure 6.7. Classified 3D model of the Sarcophagus of the Spouses.

The rendering of the classification results of the Sarcophagus is available at <https://youtu.be/oLJE8mjduo>.

6.3 BARTOCCINI'S TOMB, TARQUINIA (1)

Tarquinia was one of the most ancient cities of the Etruscan civilization. The necropolis, situated in the areas of Monterozzi and Calvario, is composed of some 6000 tombs, 60 of which decorated with paintings. The Bartoccini tomb, dated to around the 4th century BC; was discovered in 1959. It has four rooms - a central one (ca 5 m by 4 m) and three later rooms (ca 3 m by 3 m) - all connected through small corridors. The height of the tomb rooms does not exceed 3 m, and it is all painted with a reddish colour and various figures.

Combined TOF scanning and panoramic photographic surveys were carried out to get the complete 3D model of the tomb (3 million of triangles) (Fernández-Palacios et al., 2013) (Figure 6.8). The TOF range data were used to derive the geometry of the tomb, while the panoramic images to get the photo-realistic high-resolution texture.



Figure 6.8. 3D model and panoramic pictures of the Bartoccini's tomb in Tarquinia.

Part 3. Case studies

As over the centuries, the tomb has suffered from erosion caused by various reasons such as infiltration, seasoning, ageing, etc. the classification aimed to identify and quantify the deteriorated surfaces on the painted walls. Manually performing this operation would be a costly and time-consuming process.

The aim of the classification was to review the accuracy of an unsupervised clustering approach, specifically for the identification of the damaged surfaces instead of using a supervised approach.

To facilitate the clustering segmentation, and thus achieve better results, the images were initially converted from RGB to Lab* colour space (Figure 6.9).

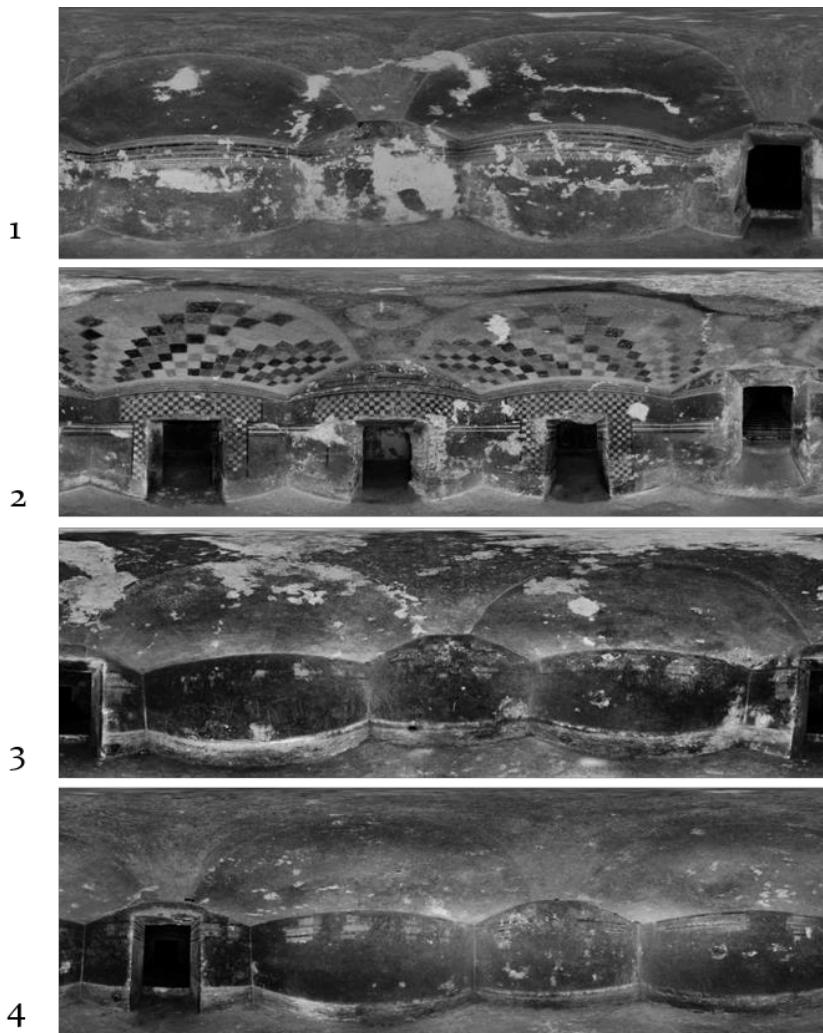


Figure 6.9. Panoramic pictures converted to Lab* colour spaces to better highlight the deteriorated areas.

Then a K-means clustering algorithm (Section 4.4) was performed to generate a pixel-based segmentation of the panoramic images (Figure 6.10). As stated in Section 2.2.3, when using the K-means clustering algorithm, the operators must specify the parameter K , that represents the total number of clusters that they want to use. For this case study, trying an incremental number of K , was possible to highlight some deteriorated areas with a $k=5$ (dark blue regions).

Chapter 6. Texture-based classification results

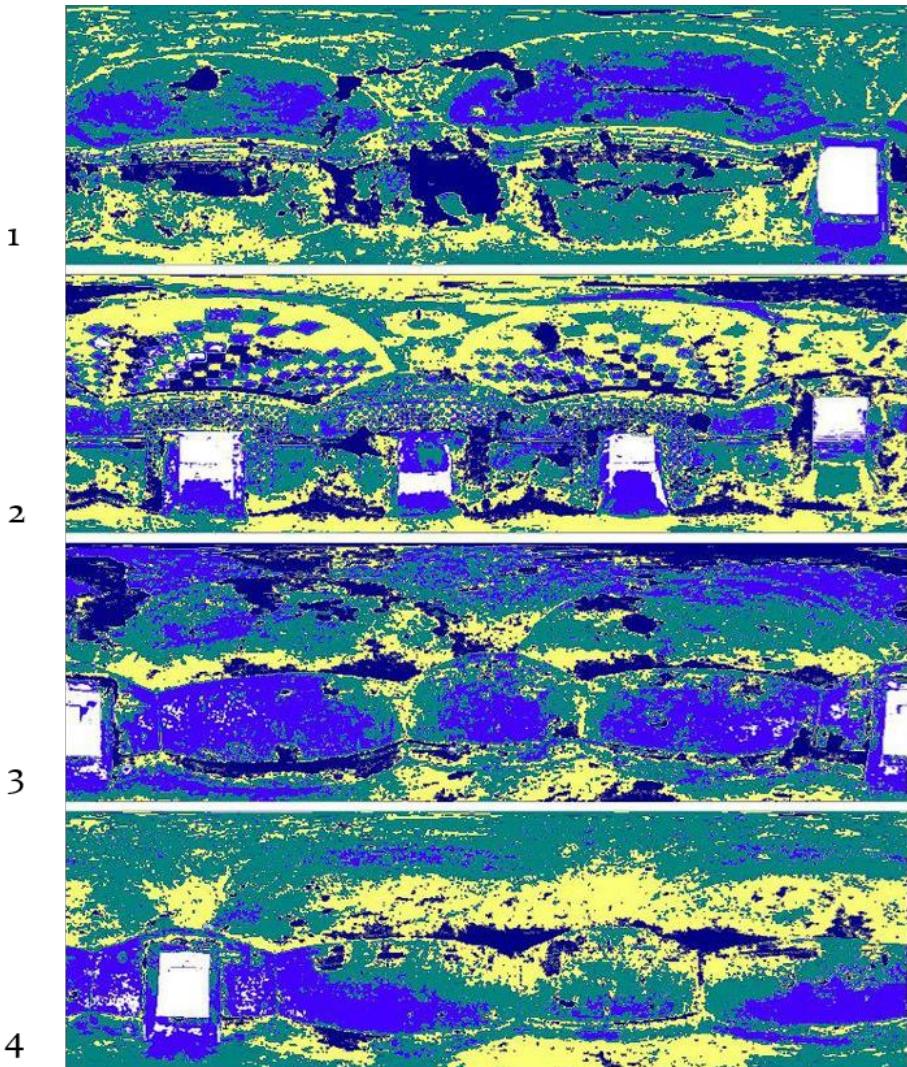


Figure 6.10. K-means clustering results with $K=5$; eroded areas in dark blue.

After analysing the results (Figure 6.10), it was clear that relying only on texture, the chromatics similarity between the deteriorated surfaces and the wall paintings caused some misclassification problems. While for *Room 1* and *Room 3* it was possible identifying the deteriorated parts, in *Room 2* and *Room 4* there were many classification errors.

The texture-based results for *Room 1* (Figure 6.11) were satisfying re-projected onto the 3D model (Figure 6.12).

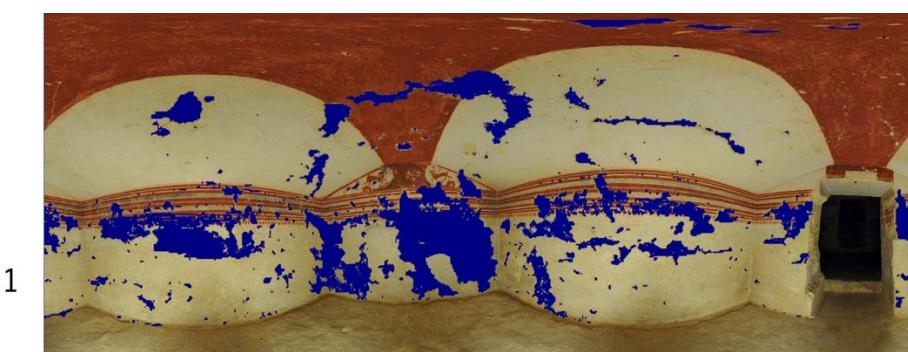


Figure 6.11. Automatically detected deteriorated surfaces.

Part 3. Case studies

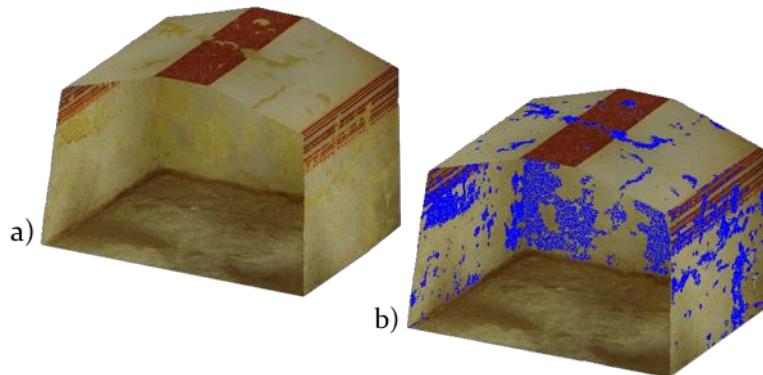


Figure 6.12. Projection onto the 3D model (**a**) of the classification results (**b**).
Bartoccini's tomb, Room 1.

A portion of the panoramic image was considered as an evaluation set (Figure 6.13a), manually classified (Figure 6.13c) and compared with the automatic prediction (Figure 6.13d), to evaluate the accuracy of the results numerically.

The overall accuracy achieved, calculated as the ratio between the number of correctly classified pixel and the total number of pixels, was of 91.15%.

It is important to say that, while the automatic segmentation with clustering took about 2 minutes to classify all the four rooms, the manual annotation (Figure 6.13b) was time-consuming (more than 1 hour for such a small portion).

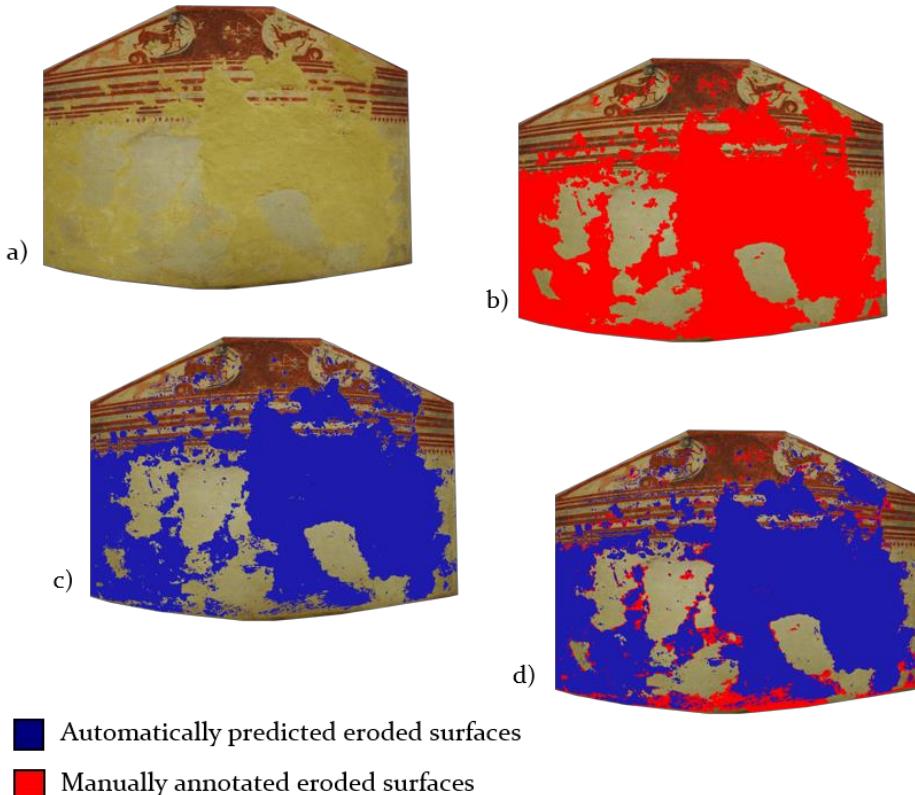


Figure 6.13. A wall of the Bartoccini's tomb (**a**) with manually annotated eroded surfaces (**b**), and automatically identified (**c**); overlap of the two results (**d**).

6.4 PORTICOES IN BOLOGNA (1)

The old porticoes of Bologna (Figure 6.14) were built during the 11th-20th centuries and can be regarded as unique from an architectural viewpoint in terms of their authenticity and integrity. Thanks to their great extension, they span approximately 40 km, and historically, the porticoes are considered a high-quality architectural work. Such structures combine various geometric shapes, different materials, and many architectural details such as mouldings and ornaments.



Figure 6.14. Porticoes in Bologna, Italy.

Become a distinctive building feature of the city, 25% of the porticoes were digitized using terrestrial photogrammetry under a project for the candidature of the porticoes as UNESCO “world heritage site”.

Considering the availability of this vast dataset (Remondino et al., 2016), different classification experiments have been conducted on this case study. In this section, the texture-based approach is applied to a small portion of the dataset (ca. 8m L' 13m H' 5 m D) (Figure 6.15a), because guiding the unwrap of a bigger part would have been a really challenging operation. A more consistent portion of the Bologna dataset is treated with a geometry-based approach in Chapter 7. In both cases, the classification was aimed at identifying the principal parts and architectural elements of the porticoes building in all cases.

Back to the texture-based approach: once generated, the UV map was first manually annotated into 10 different classes of interest (Figure 6.15b), then enriched with the radiometric features embedded in Image J Fiji (Schindelin et al., 2012). At that point, the RF model was trained to predict the classification on the entire texture (Figure 6.15c). Finally, the classified UV was re-projected onto the 3D model (Figure 6.16).

While some classes were well classified, several errors were found in those parts of the model where the texture is not homogeneous because of plaster decays. One solution might be creating additional categories for the different decays or uniform the areas that have small holes, as post-processing.

Part 3. Case studies

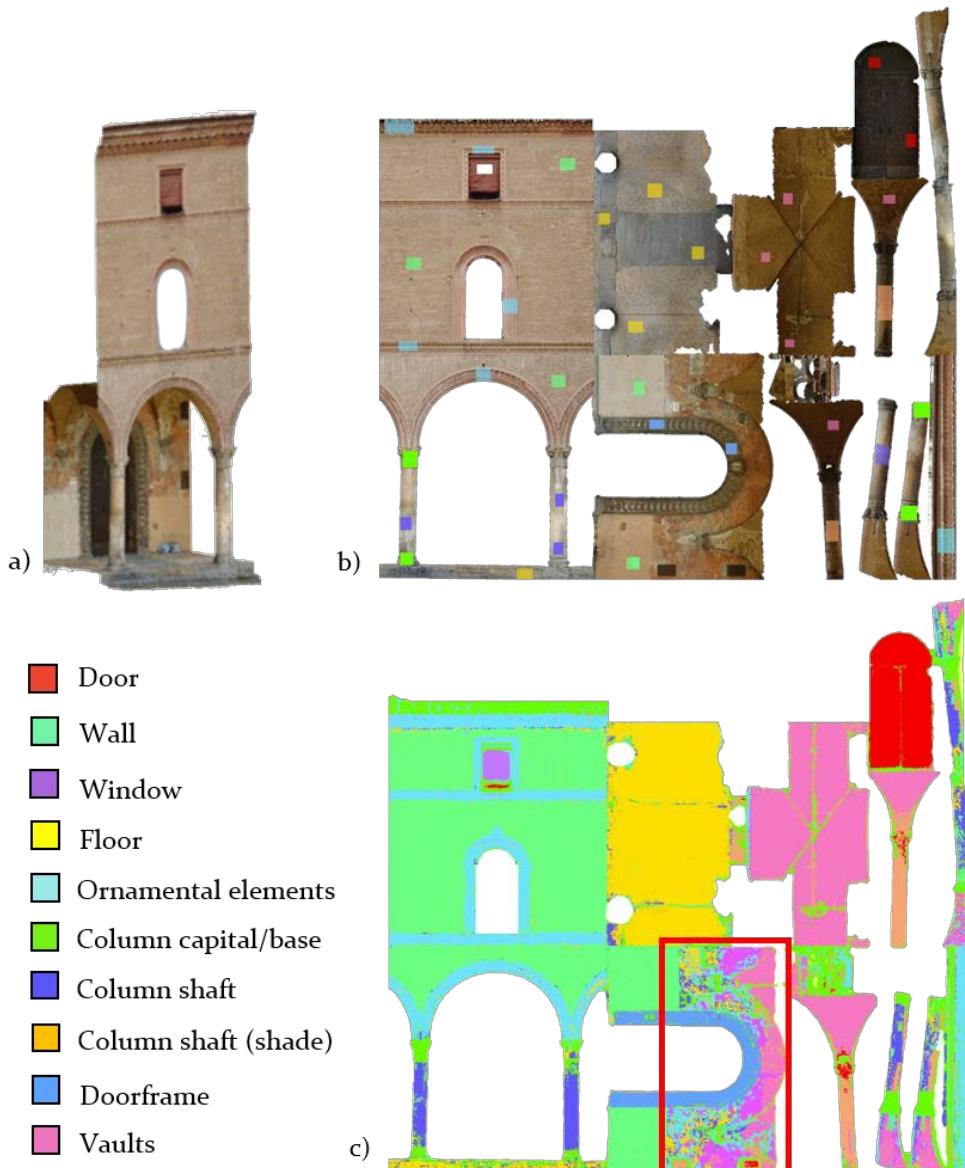


Figure 6.15. Considered portion of porticoes (a); related UV map with manual annotation (a); classified UV map (b).

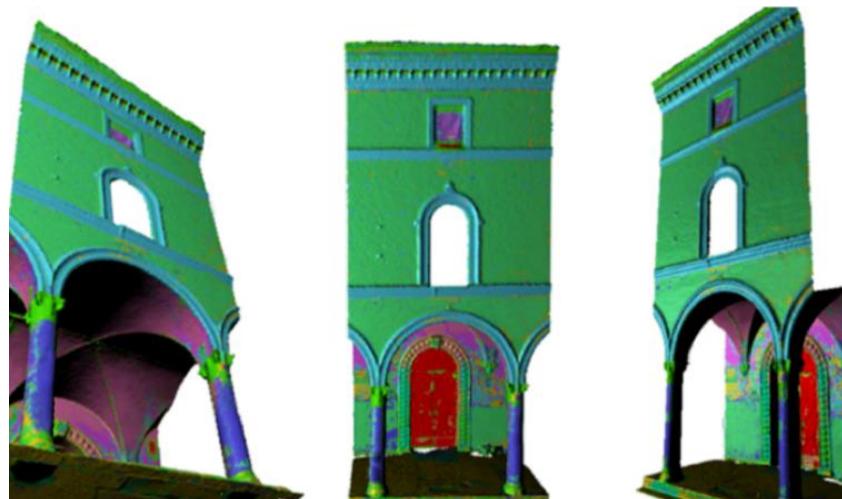


Figure 6.16. Classified 3D of a small portion of porticoes in Bologna.

Chapter 6. Texture-based classification results

REFERENCES

- Fernández-Palacios, B.J., Morabito, D. and Remondino, F., 2017. Access to complex reality-based 3D models using virtual reality solutions. *Journal of cultural heritage*, 23, pp.40-48.
- Grilli, E., Dininno, D., Marsicano, L., Petrucci, G. and Remondino, F., 2018. Supervised segmentation of 3D cultural heritage. In *2018 3rd Digital Heritage International Congress (DigitalHERITAGE) held jointly with 2018 24th International Conference on Virtual Systems & Multimedia (VSMM 2018)*, pp. 1-8. IEEE.
- Menna, F., Nocerino, E., Remondino, F., Dellepiane, M., Callieri, M. and Scopigno, R., 2016. 3D digitization of an heritage masterpiece - a critical analysis on quality assessment. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 41.
- Remondino, F., Gaiani, M., Apollonio, F., Ballaben, A., Ballaben, M. and Morabito, D., 2016. 3D documentation of 40 kilometers of historical porticoes - the challenge. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 41.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B. and Tinevez, J.Y., 2012. Fiji: an open-source platform for biological-image analysis. *Nature methods*, 9(7), p.676.
- Witten, I.H., Frank, E., Hall, M.A. and Pal, C.J., 2016. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

CHAPTER 7

Geometry-based classification results

Different heritage monuments were considered to evaluate the aforementioned geometry-based method (Chapter 5):

- The Bartoccini's tomb in Tarquinia: the same dataset used for the texture-based experiments (Section 6.3) was classified using CGAL (Giraudot and Lafarge, 2019) (Section 7.1).
- The Basilica in Paestum: it spans ca 24,5 x 54 m, and it does not exceed 10 metres in height. On the available laser scanning dataset (Fiorillo et al., 2013) was carried out a detailed analysis of the geometric features (Section 7.2). The results obtained from the analysis of the features have provided us with a key to reading the subsequent classification experiments (Section 7.3).
- The Temple of Neptune in Paestum: it measures ca 24,5 m L x 60 m D x 18 m H (in the highest part). The point cloud is the result of a combined UAV and terrestrial photogrammetric survey (Fiorillo et al., 2013). On this dataset more than one classifier was tested, to explore and compare the full potential of the machine and deep learning algorithms (Section 7.4).
- A portion of porticoes in Bologna (Italy): after the texture-based approach, a different and more significant portion of the Bologna dataset (ca 85 m L x 20 m H x 5m D) was used for some geometry-based experiments. The same dataset was treated with the features embedded in CGAL (Section 5.6) and the covariance ones (Section 5.2.1). As well as for the Temple dataset, more than one classifier was tested (Section 7.5).

7.1 BARTOCCINI'S TOMB, TARQUINIA (2)

As with the texture-based approach it was not possible to fully classify the Bartoccini's tomb (Section 6.3), a second approach based on CGAL (Giraudot and Lafarge, 2019) (Section 5.6) was used on the same dataset.

In order to get optimal results radiometric and geometric features were combined, as:

- relying only on texture, some classification errors came out due to the chromatics similarity between the deteriorated surfaces and the wall paintings (Section 6.3);
- the use of only geometry-based features allows to identify just the damaged areas below a certain depth variation threshold.

As a common practice, at first, the geometric features (embedded in CGAL) and the texture-based features (HSV values) were extracted. Second, few well-distributed portions of the point cloud were annotated on the point cloud. Then, a Random Forest classifier was trained to predict the classification on the entire dataset (Figure 7.1). The whole process took circa one hour.

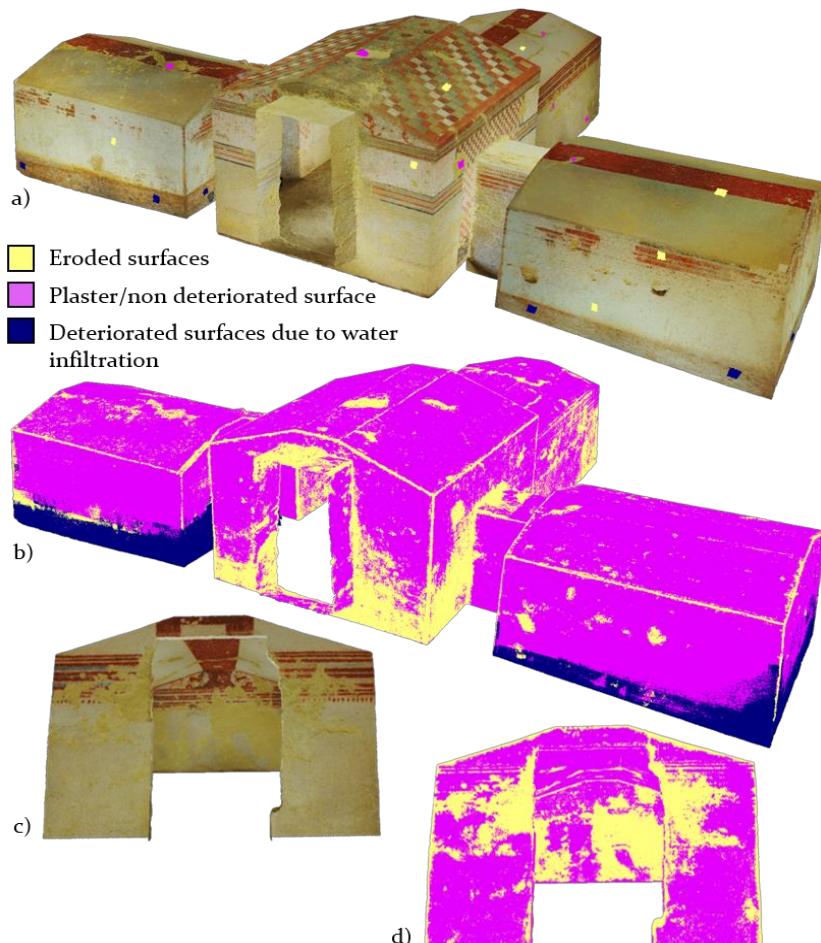


Figure 7.1. Point cloud of the Bartoccini's tomb (ca 5 million points), with manually annotated areas (a); classification result achieved using geometry- and texture-based features (b); a closer view(c)(d).

Chapter 7. Geometry-based classification results

To evaluate the results, as done for the texture-based approach (Section 6.3), a small portion of the dataset was considered, manually classified and compared with the same automatic predicted (Figure 7.2). The Overall Accuracy (13) reached in output was over 90%.

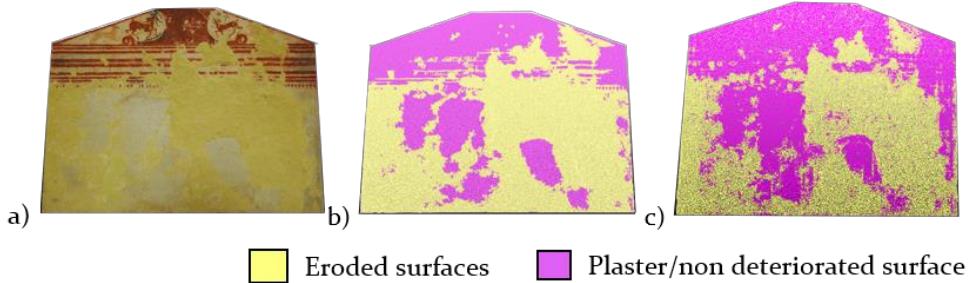


Figure 7.2. The test set for the Bartoccini's tomb (a), manually (b) and automatically (c) classified.

Given the outcome, all the eroded surfaces of the tomb were quantified. At first, the percentage occupied by each class was calculated as the ratio between the number of points belonging to the class and the total number of points of the cloud. Then, this percentage was multiplied for the total surface of the tomb, known as the sum of the triangle areas of the 3D model. In this way, it was possible to convert an automatic classification result into a real measure of deteriorated surfaces (Table 7.1). This kind of output can be beneficial for both monitoring and restoration purposes.

CLASS	Nº points	Percentage	Area (m ²)
Eroded surfaces	3539781	22.62%	42
Plaster	11014402	70.38%	131
Water infiltration	1095229	7.00%	13
TOT	15649412	100.00%	186

Table 7.1. Extracted classes and their quantification.

To better visualise the classification results, a rendering of the Bartoccini's tomb is available at <https://youtu.be/jy9PvtTcRxE>.

7.2 BASILICA, PAESTUM

Approximately ninety kilometres south of Naples, Italy, stands the ancient city of Paestum (Figure 7.3). Paestum was a major ancient Greek city on the coast of the Tyrrhenian Sea in Magna Graecia. After its foundation by Greek colonists, its name was Poseidonia, as the site was once a ceremonial centre of Poseidon (the Roman Neptune), the god of the sea. It was then conquered by the Romans who gave the city its current name Paestum. The ruins of Paestum are notable for their three Doric temples which are in an excellent state of preservation: (i) the 550 BC Temple of Hera (more known as Basilica, (ii) the 450 BC Temple of Neptune, and (iii) the smaller Temple of Ceres dated 500 BC.

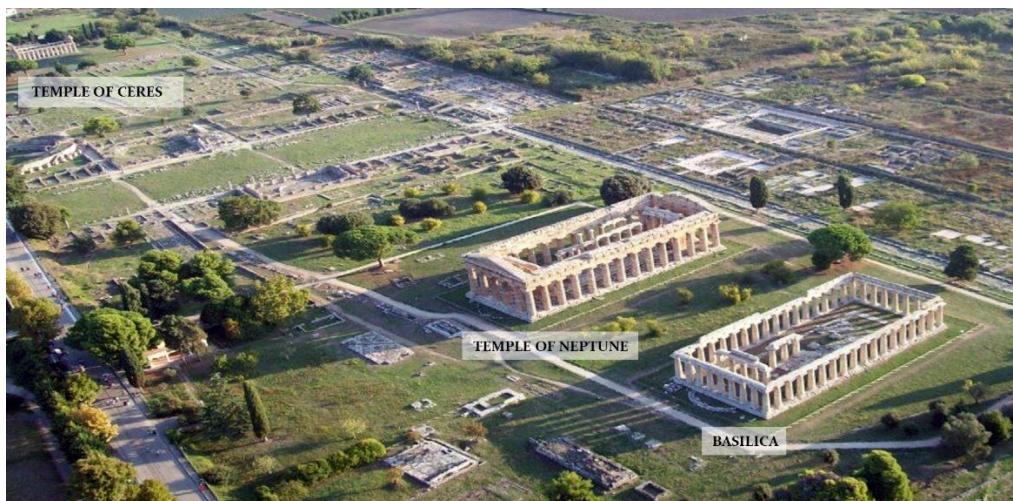


Figure 7.3. The archaeological site of Paestum, Italy.

The name "Basilica" was attributed to the Temple of Hera when it was rediscovered in the eighteenth century. As the tympanums and most of the trabeation had not survived (Figure 7.4), the functional identification as a temple of the building was uncertain. However, unlike the other temples, his function and dedication are guaranteed, thanks to inscriptions to Hera on the temple.

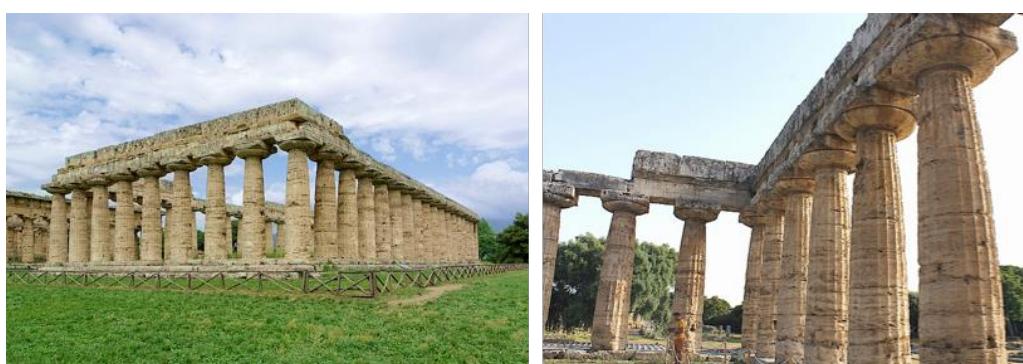


Figure 7.4. The Basilica (also called Temple of Hera) in Paestum, Italy.

In this section, and the next one, are going to be treated the laser scanning dataset of the Basilica and the photogrammetric ones of the Temple (Fiorillo et al., 2013) respectively.

7.2.1 Assessment of the feature selection workflow

In section 5.5 has been briefly introduced a workflow to optimize the feature selection (Figure 7.5). In this section is presented a more detailed explanation of the steps that have been followed, applied to the Basilica dataset. Being the architectural composition of the monument quite regular, it was enough considering a tiny portion of the point clouds as a training set on which conducting the feature experiments.

The principal aim of the workflow was to identify a small subset of features that could perform right prediction in a shorter time than using many features extracted in a conventional multi-scale approach. In particular, given the influence of the radius r in the feature response, we aimed to identify the optimal r able to discriminate our classes better.

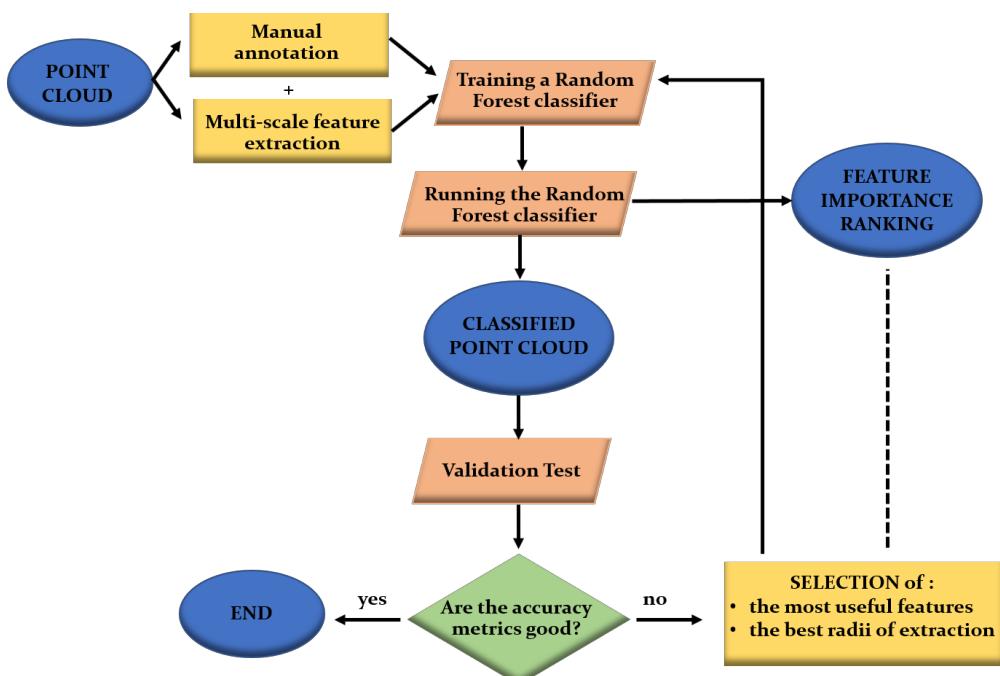


Figure 7.5. Feature selection workflow.

The first step of the flowchart consists in enriching the dataset with (i) manual annotations (8 different semantic classes) (Figure 7.6) and (ii) features (Covariance + Verticality) (Section 5.2) calculated at different search radii r (multi-scale approach). As shown in Table 7.2 the search radii have been denoted as a subscript to the name of the feature set. For example, *Planarity* (o.₂) means that the feature Planarity P_λ was calculated for a search radius of o.2 m. In order verify the different behaviours of the features connected with the architectural elements, in a first moment the features have been computed for increasing radii values between $r_{\min} = 0.2$ m and $r_{\max} = 3$ m.

Part 3. Case studies

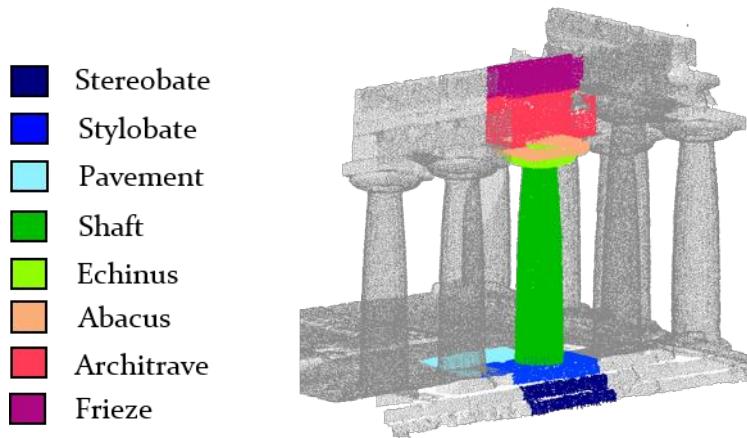


Figure 7.6. Training set for the Basilica dataset.

Once the training set was enriched with the annotated classes and the features, a Random Forest model was trained. As said in Section 5.5, the RF algorithm plays a vital role in the feature selection. Not only it provides a classification prediction on the entire dataset, but also tell us about the role the features had during the decision-making process, through feature relevance indexes (Table 7.2).

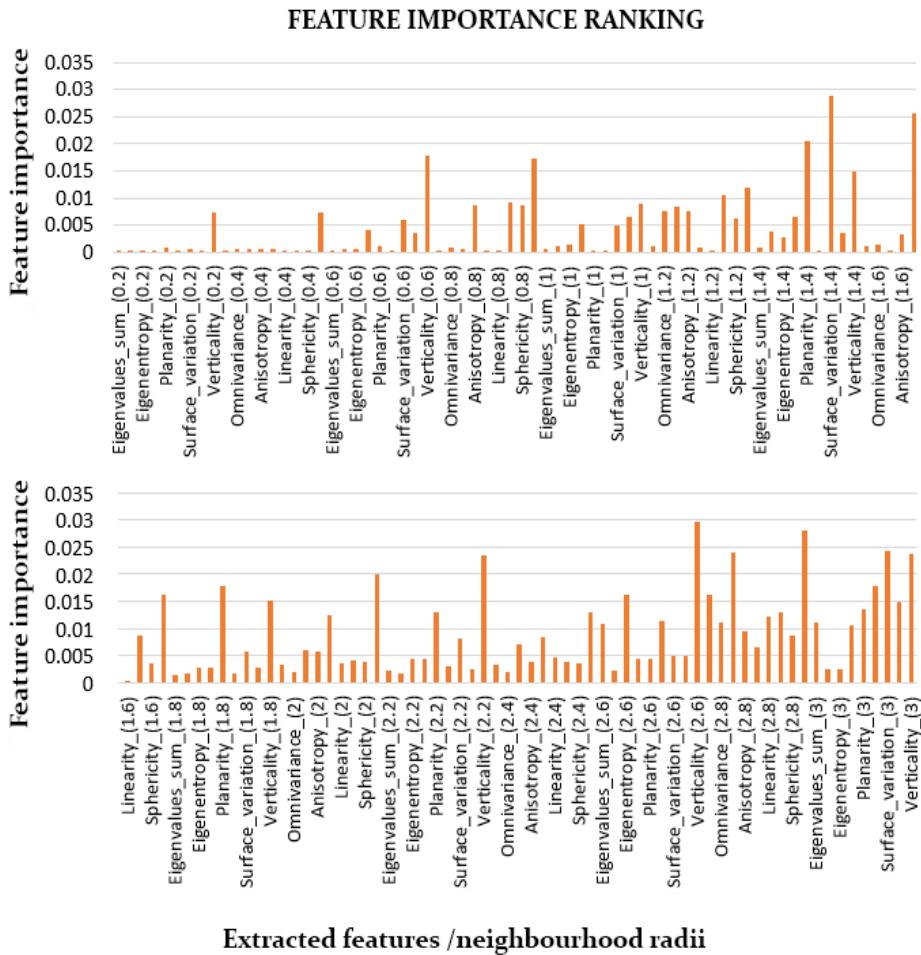


Table 7.2. Feature importance ranking for a multi-scale classification of the Basilica dataset.

Chapter 7. Geometry-based classification results

From the feature ranking (Table 7.2) we can observe that:

- the features *Eigenentropy* $E\lambda$ and Sum of eigenvalues $\Sigma\lambda$, are always among the least relevant features;
- among the covariance features, the most relevant are *Surface Variation* C_λ , *Planarity* P_λ , and *Sphericity* S_λ ;
- there is an apparent relationship between the most relevant radii of extraction of the features and the sizes of the columns; i.e. the diameter \emptyset of the Basilica's columns measures 1.4 m, and we can notice peaks in the graph around 0.7 m, 1.4 m, and 2.8m radius;
- the feature *Verticality* V is highly relevant, even when considering different neighbourhood radii; contrary with the covariance features (Section 5.5.1), its behaviour carries on similar with different radii (Figure 7.7).

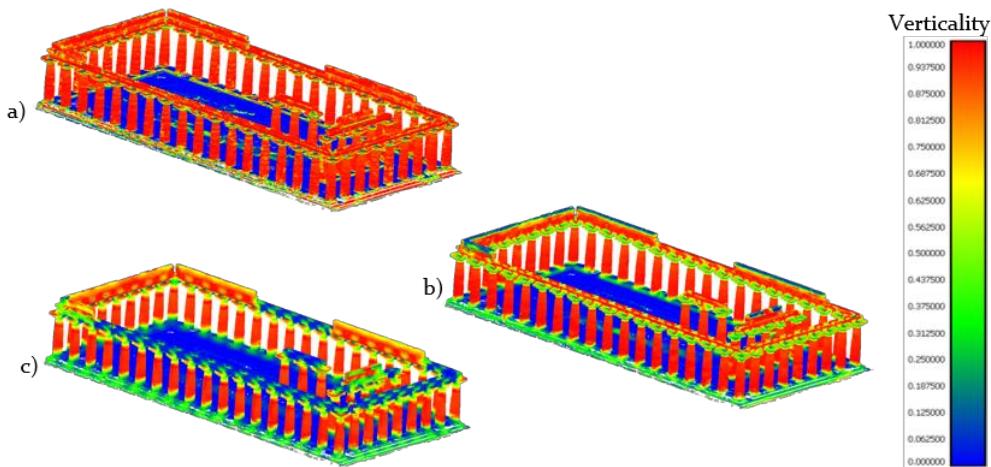


Figure 7.7. Similar behaviours of the feature *Verticality* when computed at different radius sizes: 0.4 m (a), 1 m (b), 1.4 m (c).

Next, to verify the above observations, we extracted directly 8 features, selected ad hoc and related to the dimension of the columns (radius and diameter):

- *Verticality* (0.4 m) (1 m);
- *Linearity* (0.7 m);
- *Omnivariance* (0.7 m);
- *Surface Variation* (0.7 m);
- *Anisotropy* (1.4 m);
- *Planarity* (1.4 m);
- *Sphericity* (1.4 m).

Observing the results (Figure 7.8) seems that a relationship between the covariance features (Section 5.2.1), the neighbourhood size of extractions of them, and the architectural elements was found. We can see for example, that the feature *Planarity* extracted with $r = \text{column diameter } \emptyset$ is able to highlight

Part 3. Case studies

the shafts of the columns, while the *Anisotropy* is useful to identify the capitals. The *Surface Variation* and the *Sphericity* have similar behaviours if extracted respectively at $r = \text{column ray}$, and $r = \text{column diameter } \emptyset$; in both cases, they can emphasise not only the column shafts but also separate frieze and architrave.

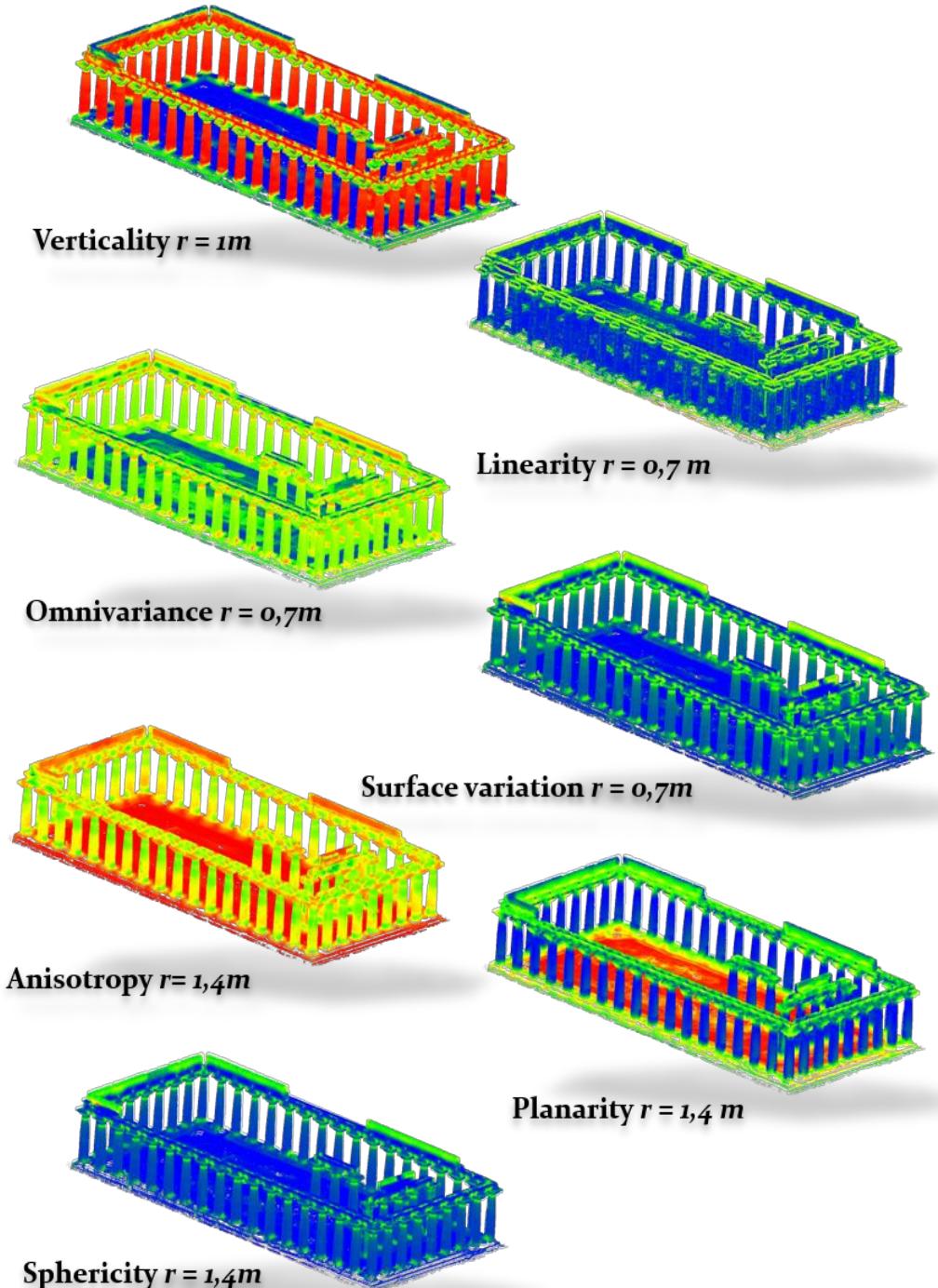


Figure 7.8. Geometric features selected ad hoc after the analysis of the feature ranking (Table 7.1).

An hypothesis about this type of results could be that the strict proportional rules and dimensions used in the construction of the Greek temples (widely studied in the architecture treatises by Vitruvio, Leon Battista Alberti, Jacopo

Chapter 7. Geometry-based classification results

Barozzi da Vignola, Palladio, etc.) (e.g., Figure 7.9) helped us to highlight the different architectural elements using just a few well-chosen scale of features.

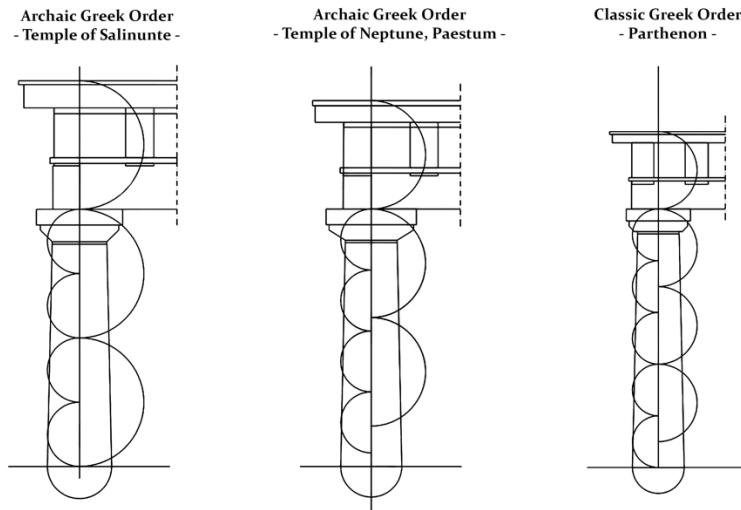


Figure 7.9. Schemes of the Greek Doric order.

Once selected ad hoc a few features, the Random Forest classifier was trained again, in order to investigate their effectiveness as a support for the classification. In parallel, the dataset was classified using other different combination of features, starting from the 135 ones extracted above with a multi-scale approach, and iteratively discarding the least important ones for the algorithm, down to seven features.

Finally, all the different classification results have been compared with respect to F1-score and time for training (Table 3).

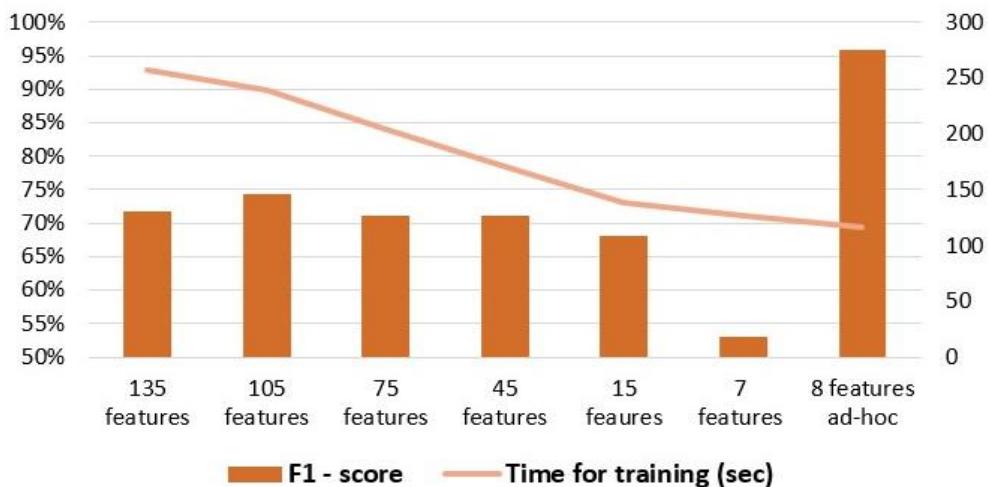


Table 7.3. Comparison of different feature combination for the Basilica classification, whit respect to F1-score and time of training.

As we can see, by iteratively discarding the least relevant features the F1-score increases at the first step then slowly decreases. In contrast, the performance of the classifier is considerably improved by selecting only a small subset of useful features. Besides, it saves processing time concerning feature extraction, training

Part 3. Case studies

time and classification. After the manual annotation (5 min), the extraction of ad hoc features (15 min), the Random Forest took ca half a minute to classify the entire datasets (500 mil points) (Figure 7.10). Results in terms of Precision, Recall, and F1-score are summarised in Table 7.4.

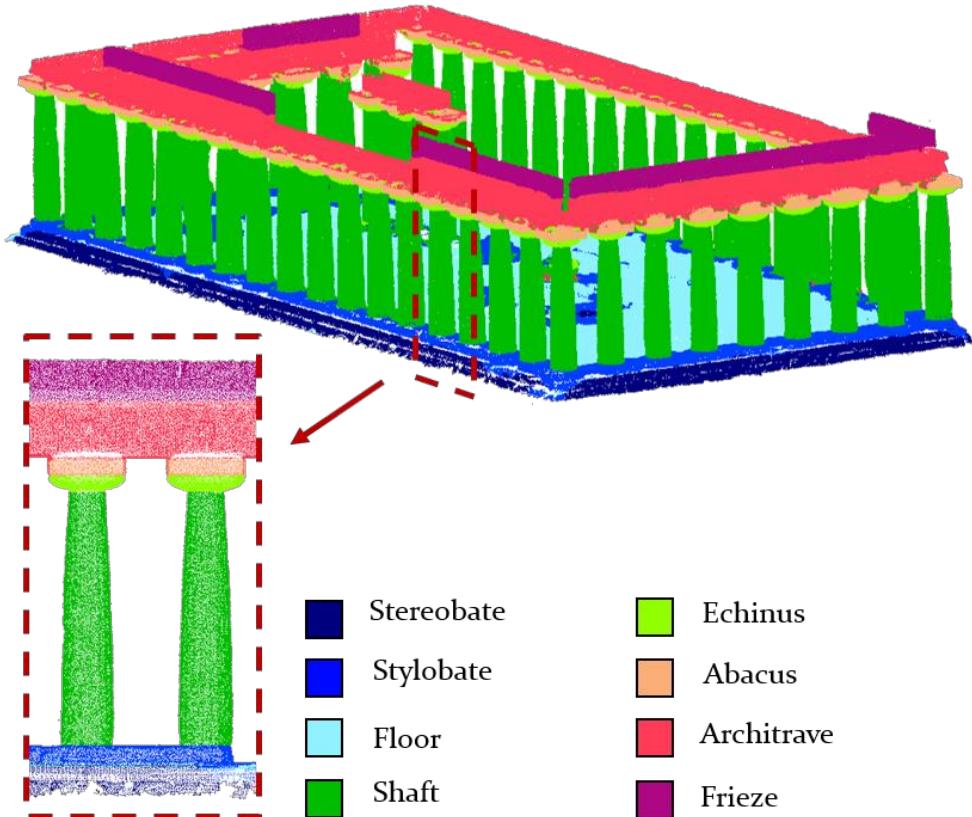


Figure 7.10. Classification result of the Basilica point cloud.

CLASS	Ster.	Stylob.	Floor	Shaft	Ech.	Abac.	Archit.	Frieze	Prec.	Recall	F1
Stereob	3325	427	13	0	0	0	0	0	88.31%	98.52%	93.14%
Stylob	50	7974	312	7	0	0	0	0	95.58%	83.99%	89.41%
Floor	0	1093	11992	0	0	0	0	0	91.65%	97.36%	94.42%
Shaft	0	0	0	48490	0	0	0	0	100 %	99.88%	99.94%
Echinu s	0	0	0	1	8390	1	7	0	99.89%	99.88%	99.89%
Abacus	0	0	0	50	10	6249	979	0	85.74%	99.98%	92.32%
Archit.	0	0	0	0	0	0	19380	0	100 %	95.13%	97.50%
Frieze	0	0	0	0	0	0	7	11406	99.94%	100%	99.97%
AVERAGE									95.14%	96.84%	95.82%

Table 7.4. Confusion matrix and accuracy metrics for the Basilica dataset, using ad hoc geometric features.

7.3. AD HOC FEATURES RELIABILITY

After the experiments on the Basilica, it was important to verify if the covariance feature extraction at neighbourhood size proportional to the column radii, was working similarly on other datasets. From Table 7.5 we can see that the hypothesis was verified and different architectural elements (e.g., shafts, capitals, friezes) were highlighted choosing ad hoc features. Given the above, the experiments in the next sessions are all based on ad hoc selected features.

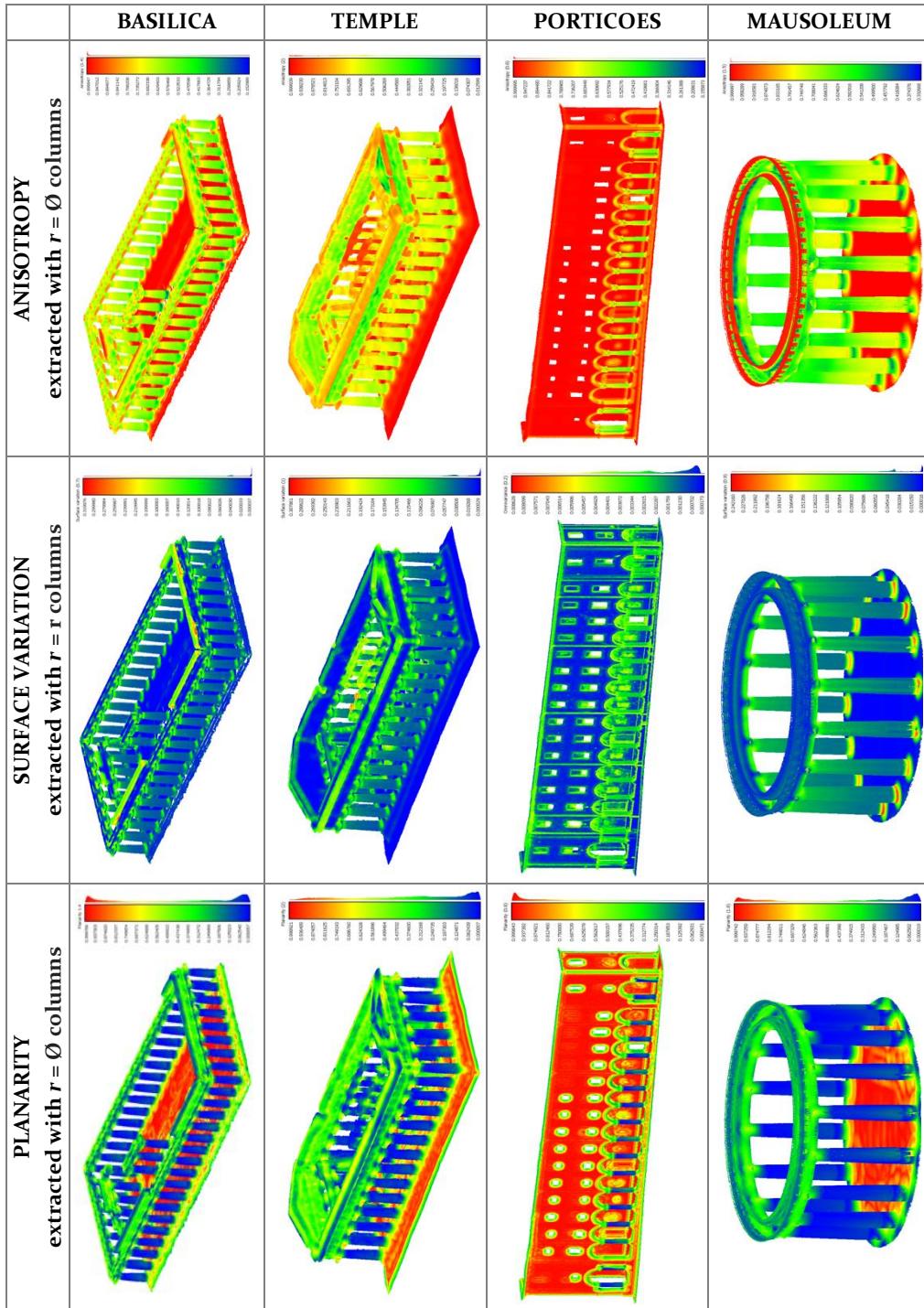


Table 7.5. Similar behaviours of the covariance features ad hoc extracted.

7.4 TEMPLE OF NEPTUNE, PAESTUM

The best preserved of the temples in Paestum is the so-called Temple of Neptune. In the eighteenth century, recognised as the biggest one, it was assumed that it must have been dedicated to the patron god of the city (Neptune). The dedication is almost certainly wrong, as the terracotta votive figurines found in the sanctuary show female types commonly identified as Hera.



Figure 7.11. Temple of Neptune in Paestum, Italy.

The Temple measures ca 24,5 x 60 m and consists of 6 frontal and 14 lateral columns while in the interior area it has two rows of double ordered columns. The available dataset is the result of a combined UAV and terrestrial photogrammetric survey (Fiorillo et al., 2013).

At first, to speed up the computational process, the point cloud was subsampled; the data on which the experiments were conducted consists of some 2.8 million points (Figure 7.12).



Figure 7.12. Photogrammetric point cloud of the Temple of Neptune in Paestum (ca 2,8 million points).

Then, to semantically classify the monument, ten different classes corresponding to the architectural elements were identified, and manually annotated (Figure 7.13). As the training set must give a key to read the entire

Chapter 7. Geometry-based classification results

structure, the annotations were distributed in different parts of the dataset, so to include all the classes and the diverse types of columns.

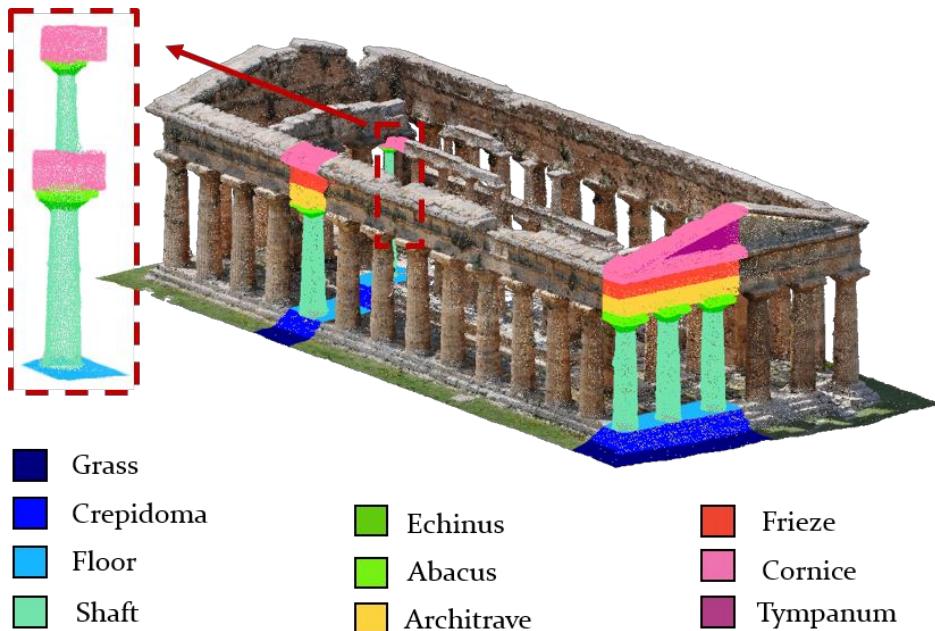


Figure 7.13. Photogrammetric point cloud of the Temple of Neptune with some manually annotated portions used for the training.

After the annotation, it was essential to extract geometric features able to highlight the architectural elements. Given the successful result of the previous experiment (Section 7.2), further verified in Section 7.3, a few covariance features were computed at neighbourhood proportional to the three different sets of columns of the Temple (diameters 0.8 m, 1.4 m, 2 m) (Figure 7.14).

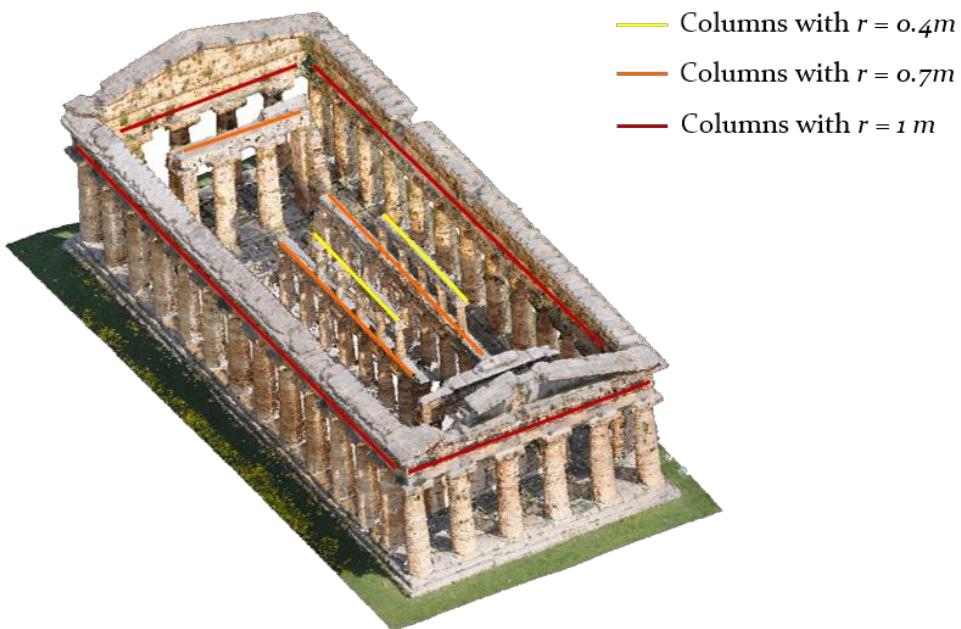


Figure 7.14. Three different sets of columns highlighted on the Temple dataset.

Part 3. Case studies

As demonstrated in Table 7.6, extracting the feature *Surface Variation* at neighbourhood sizes equal to the three sets of column radii allows detecting the respective column shafts. Similarly, with the *Anisotropy* calculated at the diameter sizes, all the different capitals are emphasised.

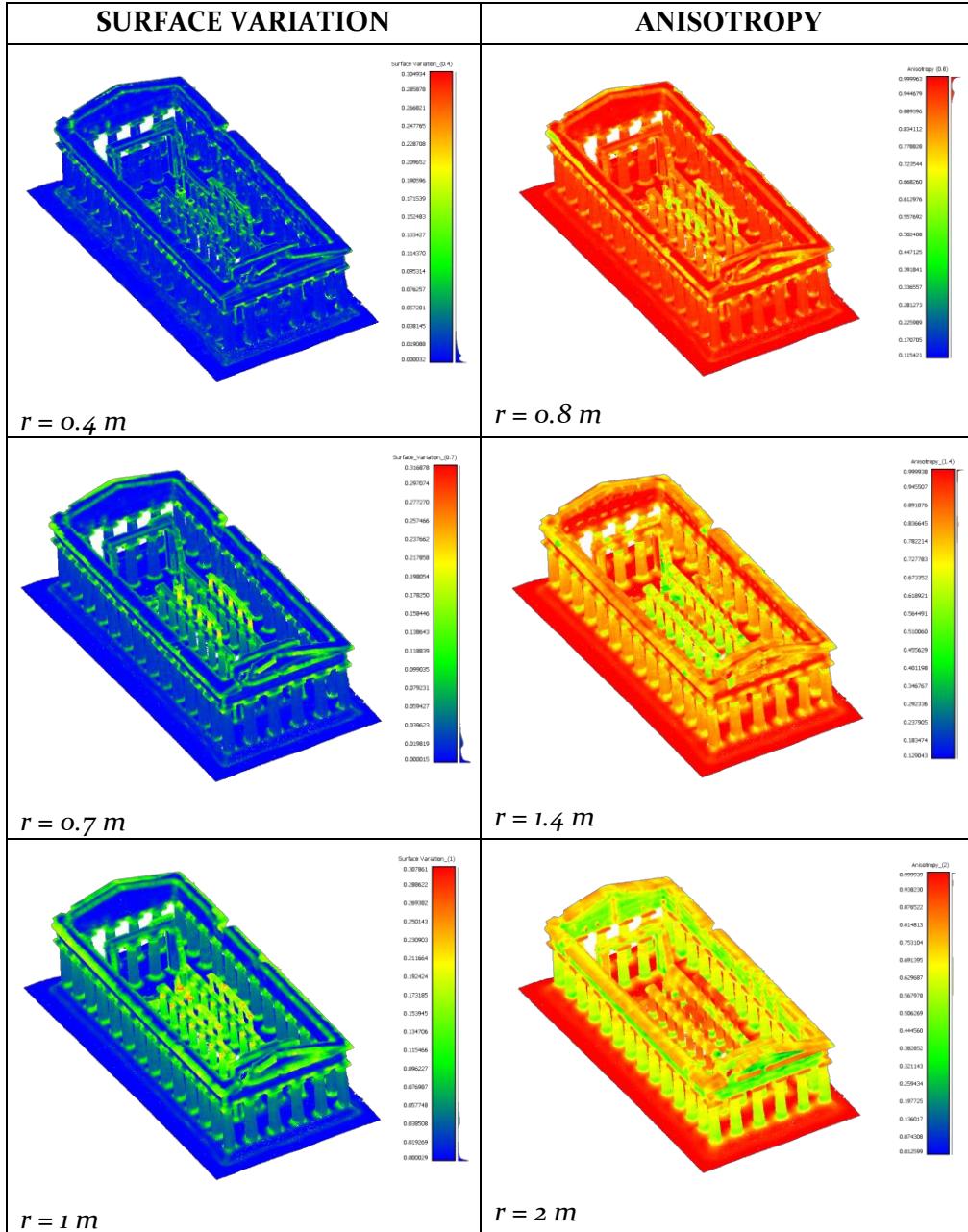


Table 7.6. Similar behaviour of the covariance features ad hoc extracted.

Once the training set was enriched with annotated classes and features, different Machine and Deep Learning models (described in Section 5.3) were trained to predict the labels on the entire dataset.

When using neural networks, the classification was carried out with and without considering the decentralised coordinates of the points (Section 5.2). The practice of using decentralised coordinate with CNN has been verified to be useful for the classification in the geospatial environment (Özdemir and

Chapter 7. Geometry-based classification results

Remondino, 2019); however, their applicability in an architectural contest had not been explored yet.

Table 7.7 summarises all the accuracy metrics reached with the different approaches. As we can observe from the diagram, the highest levels of accuracy were achieved using the machine learning approaches (RF and OvO).

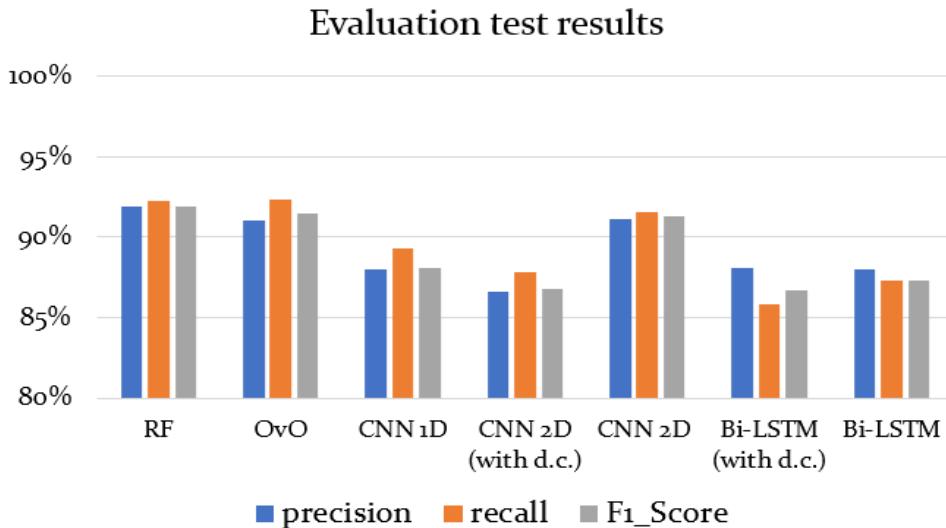


Table 7.7. Summary of the classification results for the Temple dataset achieved with different ML/DL methods.

Table 7.8 displays in parallel the per-class F1-score results for each predictive model used. The F1-score averages are between 86.69 % with Bi-LSTM and 91.92% with Random Forest.

CLASS / ALGORI-THM	F1-Score						
	RF	OvO	CNN 1D	CNN 2D (d.c.)	CNN 2D	Bi-LSTM (d.c.)	Bi-LSTM
Grass	98.52%	99.05%	99.05%	99.08%	98.89%	98.20%	97.95%
Crepid.	95.34%	95.42%	95.30%	96.22%	95.84%	92.48%	91.57%
Floor	97.98%	97.84%	97.31%	97.89%	98.02%	96.68%	96.48%
Shaft	99.04%	98.73%	98.25%	98.30%	98.77%	97.60%	97.90%
Echinus	88.45%	84.40%	81.03%	79.54%	84.91%	78.15%	76.53%
Abacus	81.66%	80.64%	74.85%	75.93%	79.23%	77.86%	72.94%
Archit.	91.16%	91.90%	78.56%	80.22%	90.69%	89.10%	89.53%
Frieze	87.31%	87.74%	77.50%	66.97%	87.31%	78.13%	86.02%
Cornice	94.55%	93.90%	93.24%	92.73%	93.54%	87.76%	89.52%
Tymp.	85.19%	84.66%	85.99%	80.55%	85.31%	70.94%	74.14%
AVERAGE	91.92%	91.43%	88.11%	86.74%	91.25%	86.69%	87.26%

Table 7.8. A summary of all the tested ML/DL models reporting per-class F1-score.

With a heritage dataset, the regular use of decentralised coordinates for DL approaches reduced the F1-Score average. However, from closer analysis, we can observe that the F1-Score decreases in particular when the classes share the same geometry (e.g., Architrave, Frieze, and Tympanum). What said above represents one of the most challenging points for heritage classification, as there's not

Part 3. Case studies

always a correspondence between shape/colours and semantics for the architectural classes.

Table 7.9 reports the results obtained with the RF classifier, including confusion matrix and accuracy metrics. Each row of the matrix represents the instances in an actual class (ground truth), while each column represents the instances in a predicted class. In general, we can observe again that most of the classification errors are between classes with similar geometric properties, such as "Abacus" and "Architrave", or also "Frieze", "Cornice" and "Tympanum (Figure 7.15).

CLASS	Grass	Crep.	Floor.	Shaft	Echinu s	Abacu s	Archit r.	Frieze	Cornic e	Tymp.	Prec.	Recall	F₁
Grass	56998	1647	o	o	o	o	o	o	o	o	97.19 %	99.88 %	98.52 %
Crepid	67	38389	484	497	o	o	o	o	o	o	97.34 %	93.43 %	95.34 %
Floor	o	958	62211	993	o	o	o	o	o	o	96.96 %	99.03 %	97.98 %
Shaft	o	54	125	169950	172	56	4	o	268	o	99.60 %	98.49 %	99.04 %
Echin.	o	o	o	966	18717	988	4	o	40	o	90.35 %	86.62 %	88.45 %
Abacus	o	o	o	63	2651	23941	3998	o	636	o	76.52 %	87.54 %	81.66 %
Archit r.	o	o	o	o	o	2140	50903	2600	1	o	91.48 %	90.84 %	91.16 %
Frieze	o	o	o	o	o	o	1036	34609	334	33	96.10 %	79.98 %	87.31 %
Cornic e	o	42	o	90	67	223	90	4569	106415	2071	93.70 %	95.41 %	94.55 %
Tymp	o	o	o	o	o	o	4	1492	3841	21393	80.03 %	91.05 %	85.19 %
AVERAGE											91.93 %	92.23 %	91.92 %

Table 7.9. RF classification results: Confusion Matrix and per-class accuracy for the Temple dataset.

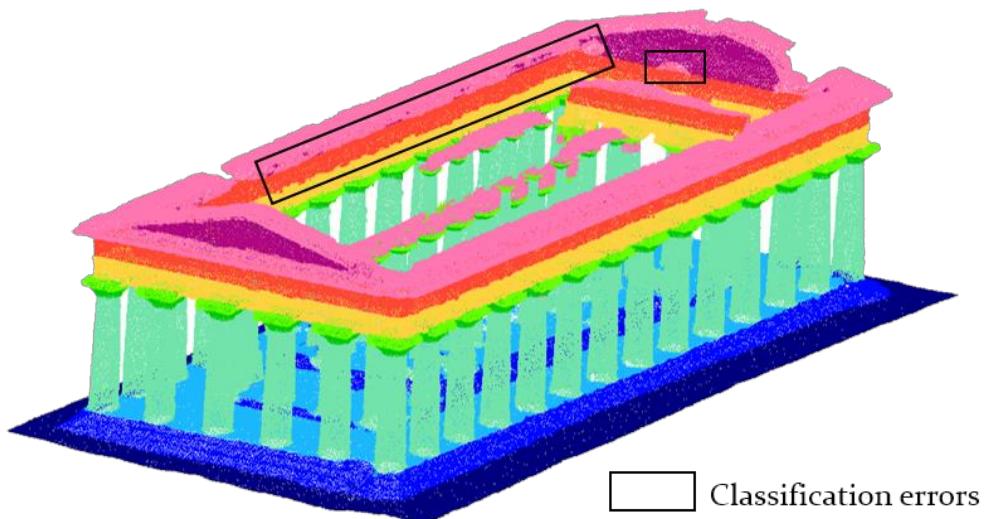


Figure 7.15. Some classification problems.

Chapter 7. Geometry-based classification results

To summarise the work done on the Temple, as seen for the Basilica dataset, the use of ad hoc covariance feature has been demonstrated to be effective for the classification results. Different classifiers were explored, and ML approaches outperformed DL methods. In terms of time expense, the manual annotation took 20 minutes, while the feature extraction around 30 minutes (20 covariance features were calculated). The DL approaches required about 30 minutes on GPU to be trained, while the ML ones completed the training in 10 minutes on a CPU. Once the models were trained, a prediction for the classification of the entire model was given in a few minutes.

Once classified, the dataset could be separated into the different semantic parts (Figure 7.16). Such a result represents a good starting point to generate a HBIM model.

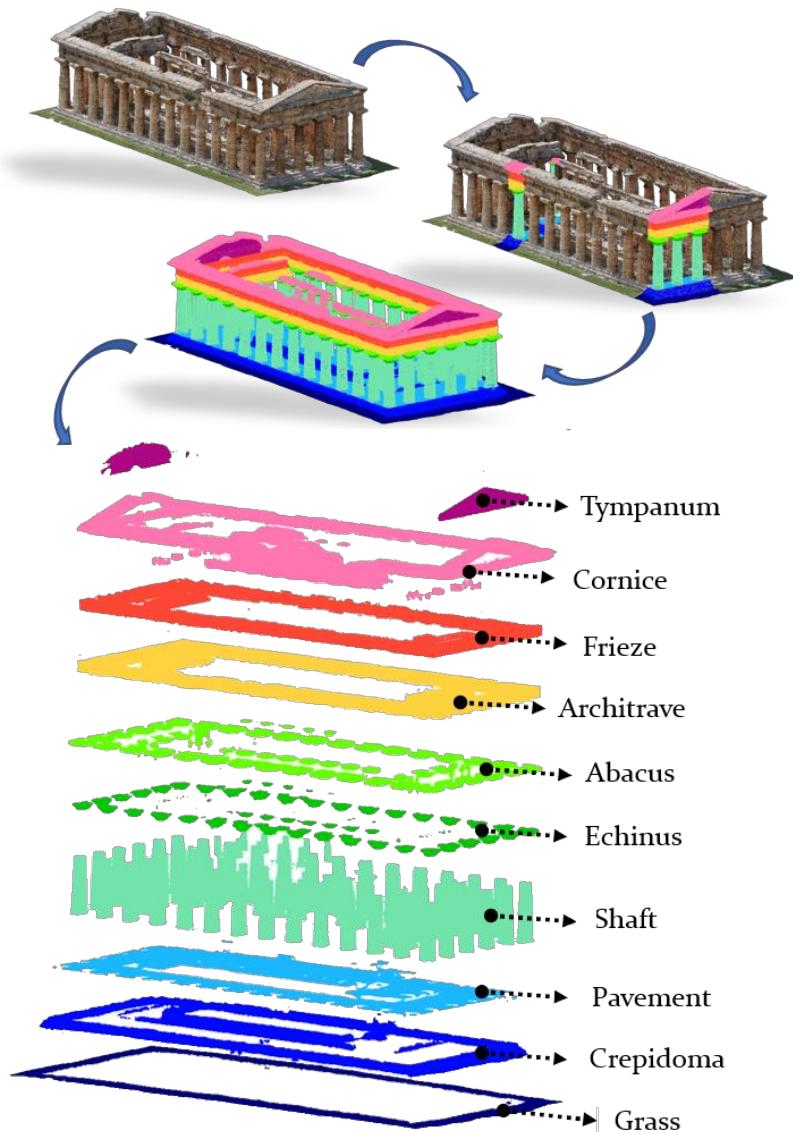


Figure 7.16. Exploded view of the Temple of Neptune dataset after the automated classification.

The rendering of the classification results of the Temple is available at <https://youtu.be/8-muH633ud8>.

7.5 PORTICOES IN BOLOGNA (2)

Given the problems we had with the texture-based approach on the Bologna dataset (Section 6.4), two different supervised approaches based on geometric features were explored on a different and bigger portion of the same dataset (Figure 7.17).



Figure 7.17. Point cloud of some porticoes in Bologna (ca 85 m L x 20 m H x 5m D), 1 million points (Remondino et al., 2016).

For both cases, the classification aimed to semantically annotate the different architectural and decorative elements of the building.

7.5.1 Classification using CGAL

At first, the CGAL method (Giraudot and Lafarge, 2019) presented in Section 5.6 and used for the classification of the Bartoccini's tomb (Section 7.1), was experimented.

The CGAL interface allows annotating the point cloud quickly by manually distributing the annotation of the classes directly on the point cloud. 14 different classes were identified for this case study (Figure 7.18). Even if it is fast, for a complex dataset like this one, such operation is not user-friendly, and the risk to include a portion of point cloud into a wrong class is high.



[Blue square]	Road	[Green square]	Capital	[Red square]	Curtain
[Blue square]	Pavement	[Green square]	Arch	[Red square]	Vertical drainpipes
[Dark Blue square]	Facade	[Yellow square]	Window / door	[Pink square]	Cornice
[Light Blue square]	Base	[Yellow square]	Vault	[Purple square]	Horizontal drainpipes
[Light Blue square]	Shaft	[Orange square]	Moulding		

Figure 7.18. Manual annotation of the classes of interest for the classification of the Bologna dataset, using CGAL interface.

After the annotation of the dataset, the predefined features provided by CGAL were extracted at 10 different scales. In addition to the geometric features, also the HSV values (Section 4.2) were considered for the classification task.

Finally, a Random Forest algorithm (ETH Zurich Random Forest Template, 2015) was defined and trained. Once generated, the classifier performed a prediction on the entire point cloud (Figure 7.19).

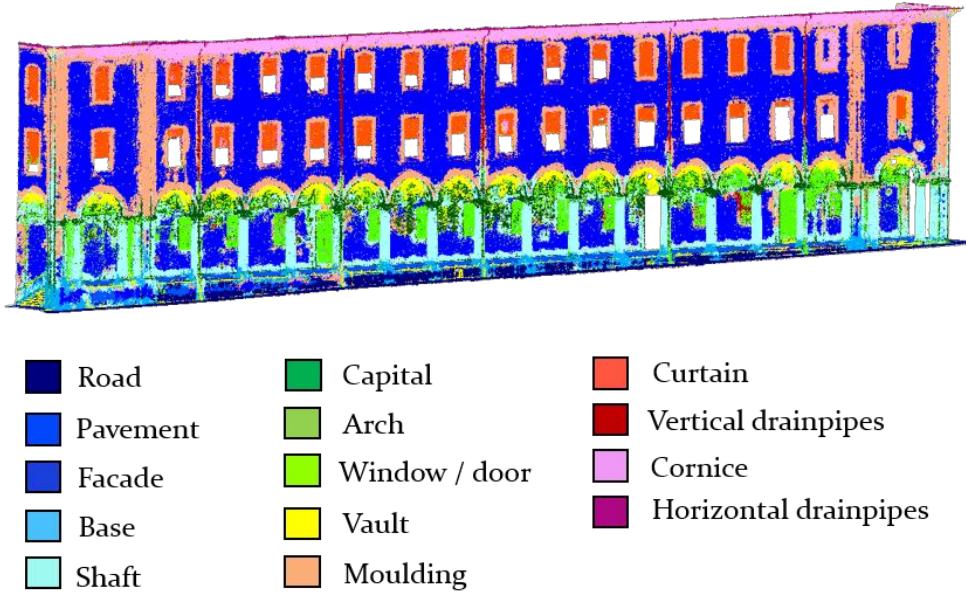


Figure 7.19. Classification results of the Bologna dataset using CGAL.

The classes *Road*, *Pavement*, *Vault*, and *Cornice* were correctly identified by CGAL, as the classifier relies on several features correlated with the z coordinates of the points (*Elevation*, *Height above*, *Height below*). Unfortunately, precisely as with the texture-based approach results, many misclassification problems came out under the porticoes where the plaster surface is not homogeneous. Probably, in this case, the colour component was preponderant over the geometric features. The same hypothesis could have caused the misclassification problems we had within the points belonging to the class *Arch*, mostly classified as *Moulding* for their colour.

7.5.2 Classification using ad hoc features

Given the classification problems with CGAL (Section 7.5.1), on the same dataset was tested the approach based on ad hoc features.

In this case, the manual annotation was controlled by the segment tool within the open-source software Cloud Compare, and it took about 20 minutes (Figure 7.20).

Part 3. Case studies

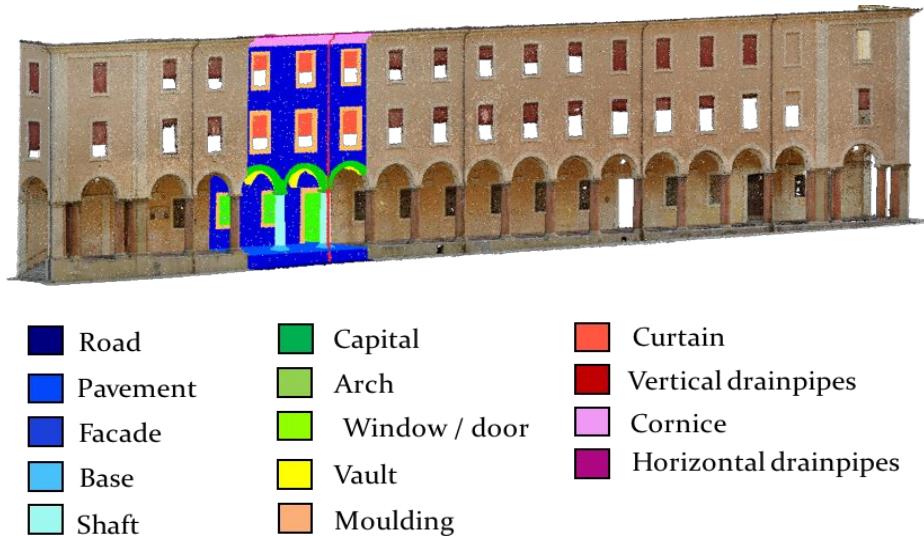


Figure 7.20. Manual annotation of the class of interest for the classification of the Bologna dataset, using CloudCompare segment tool.

After the annotation, some significant covariance features were extracted (Figure 7.21). The choice of the feature neighbourhood sizes depended on three factors:

- column size (radius = 0.4m) → the features *Planarity*, *Anisotropy*, and *Sphericity* extracted at the diameter measure could highlight the different components of the columns (bases, shafts, capitals);
- facade decorations → to detect mouldings and drainpipes the feature *Omnivariance* was calculated at a small neighbourhood size;
- intercolumniation → the spacing between columns (4 m) was measured to bring to light the category *Arch*, using the *Surface Variation* computed at 2 m.

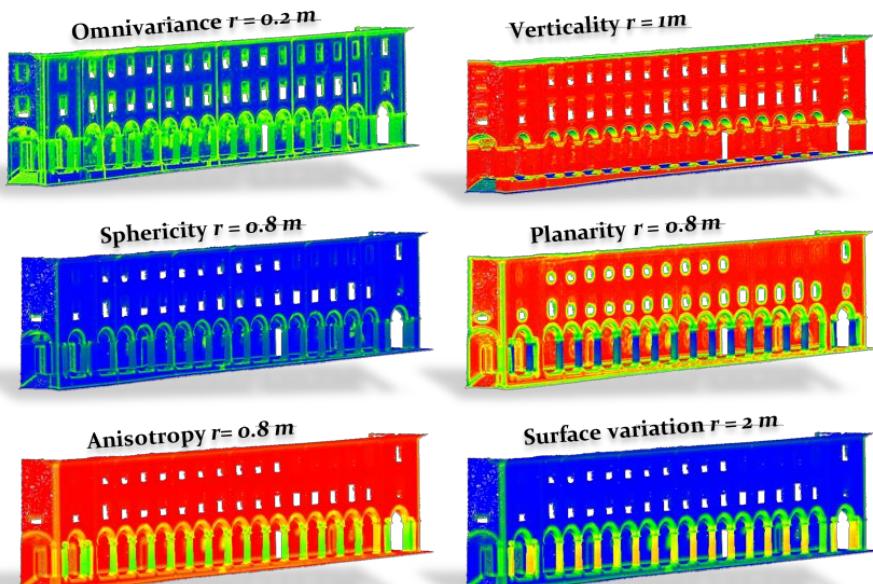


Figure 7.21. Significant geometric features highlighting the architectural elements of the Bologna dataset.

Chapter 7. Geometry-based classification results

Together with the geometric features, also the RGB values of the point cloud were given in input to train the predictive models. In fact, without using the colour components, it would have been hard distinguishing categories such as *Curtains*, *Wall* and *Window*, that present similar geometric properties.

As well as in the previous section, also for the Bologna dataset, many predictive models have been trained. Table 7.10 summarises the different accuracy metrics achieved, while in Table 7.11 the per-class F1-score results are displayed.

Evaluation test results

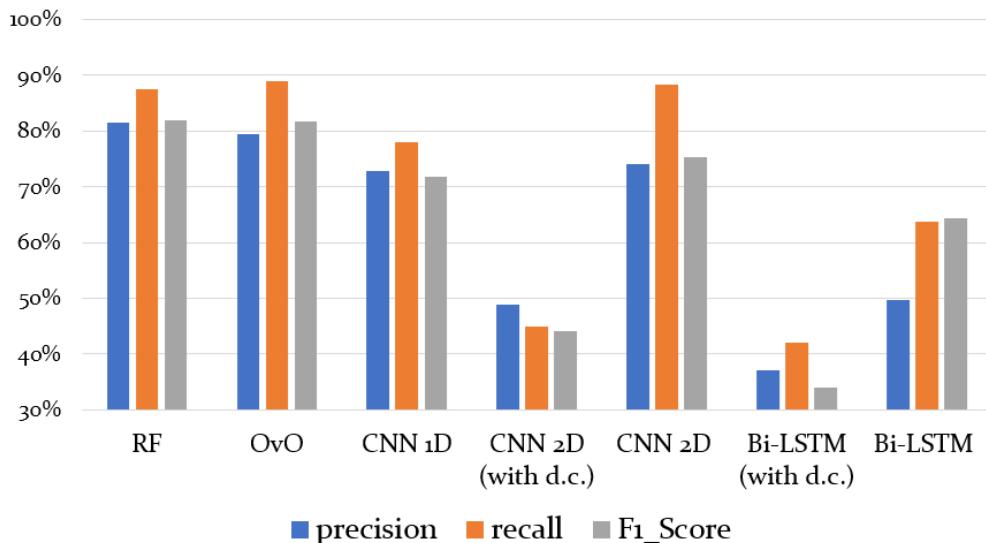


Table 7.10. Summary of the classification results for the Bologna dataset achieved with different ML/DL methods.

CLASS / ALGORIT.	F1-Score						
	RF	OvO	CNN 1D	CNN 2D (with d.c.)	CNN 2D	Bi-LSTM (with d.c.)	Bi-LSTM
Road	35.28%	46.54%	87.29%	0.00%	2.93%	3.40%	9.98%
Facade	92.05%	90.79%	88.61%	81.41%	91.35%	78.70%	83.68%
Pavement	81.59%	86.48%	89.20%	71.14%	80.47%	75.12%	78.50%
Base	82.85%	86.58%	65.96%	0.00%	83.48%	0.00%	0.00%
Shaft	97.17%	96.90%	93.23%	82.10%	97.90%	69.95%	82.28%
Capital	80.60%	82.90%	58.52%	0.00%	64.58%	0.00%	0.00%
Arch	88.86%	86.90%	67.05%	40.01%	88.48%	21.58%	53.06%
Wind/door	80.89%	79.90%	61.48%	65.67%	77.52%	53.15%	68.97%
Vault	93.94%	92.90%	88.82%	84.33%	94.57%	80.30%	85.99%
Molding	70.88%	58.77%	48.14%	23.02%	66.26%	0.02%	0.43%
Curtain	92.60%	89.11%	81.15%	82.98%	95.71%	6.21%	90.02%
Drainp. V.	68.27%	62.63%	21.82%	0.00%	57.32%	0.00%	0.00%
Cornice	96.26%	95.78%	85.84%	87.32%	94.23%	86.57%	90.42%
Drainp. H.	85.34%	88.02%	69.26%	0.00%	60.29%	0.00%	0.00%
AVERAGE	81.90%	81.73%	71.88%	44.14%	75.37%	33.93%	45.95%

Table 7.11. A summary of all tested ML/DL classification methods reporting the per-class F1-score for the Bologna dataset.

Part 3. Case studies

Again, the best-succeeded results were related to the machine learning approaches. However, while in the previous case study the lower value of F1-score was the 86.69% achieved with Bi-LSTM, in this case, there was a suspicious change between the lower 33,93% with Bi-LSTM and the higher 81,90% with RF. Even if the features were handcrafted for the case study, the class structure complexity of the dataset caused lower accuracy metrics. Moreover, it seems that the DL approach was not suitable for this kind of dataset.

To better understand the classification errors, we can have a look at the following confusion matrix related to the Random Forest prediction (Table 7.12). In red are highlighted the lower values of accuracy, found for the classes *Road*, *Moulding* and *Horizontal drainpipes*.

Most of the points belonging to the class *Road* were misclassified as *Pavement*; this problem can be easily solved, adding to the training data some extra-features related to the height of the points. We have seen that the CGAL approach was entirely distinguishing the two categories. If we did not consider the accuracy values relative to the class *Road*, the average of the results would increase up to 87%.

Having particular attention to the False Negative values concerning the classes *Moulding* and *Horizontal drainpipes*, we can see that points belonging to these two classes were often attributed to other classes with similar geometries (e.g. *Facade*, *Window/door*).

CLASS	Road	Facad e	Pav.	Base	Shaft	Cap.	Arch.	Win/ door	Vault	Moul ding	Curt.	Drain v.	Corn.	Draip H.	Prec.	Rec.	F1
Road	1411	43	5105	0	0	0	0	0	0	0	0	30	0	0	21.41 %	100 %	35.28 %
Facade	0	46838	1968	73	0	0	145	95	254	660	62	151	14	0	93.19 %	90.93 %	92.05 %
Pav.	0	59	16695	282	0	0	0	21	1	1	0	25	0	0	97.72 %	70.03 %	81.59 %
Base	0	8	17	1604	226	0	1	11	0	1	0	3	0	0	85.73 %	80.16 %	82.85 %
Shaft	0	0	0	0	9143	0	0	0	0	0	0	9	0	0	99.9 %	94.58 %	97.17 %
Capital	0	0	0	0	246	1101	178	0	73	0	0	22	0	0	67.9 %	99.01 %	80.6 %
Arch	0	169	0	0	0	0	6179	0	757	5	0	0	0	0	86.91 %	90.91 %	88.86 %
Win/do or	0	27	0	0	0	0	0	2321	0	242	0	0	0	0	89.61 %	73.71 %	80.89 %
Vault	0	1060	0	0	0	1	218	0	18331	0	0	0	0	0	93.48 %	94.41 %	93.94 %
Mouldi ng	0	2919	55	19	0	0	7	693	0	6526	205	3	0	0	62.59 %	81.71 %	70.88 %
Curtain	0	49	0	0	0	0	0	0	198	3464	4	0	0	0	93.24 %	91.96 %	92.60 %
Drainp. V.	0	221	0	23	52	10	69	8	0	354	36	1150	47	0	58.38 %	82.20 %	68.27 %
Cornice	0	118	0	0	0	0	0	0	0	0	0	2	4746	175	94.15 %	98.46 %	96.26 %
Drainp. H.	0	0	0	0	0	0	0	0	0	0	0	0	13	547	97.68 %	75.76 %	85.34 %
AVERAGE															81.57 %	87.42 %	81.90 %

Table 7.12. RF classification results: Confusion Matrix and per-class accuracy for the Bologna dataset. In orange and red are highlighted some anomalies.

Chapter 7. Geometry-based classification results

To classify the entire dataset (Figure 7.22) it was enough less than one hour using a RF model. The manual annotation was undoubtedly the most time-consuming part of the classification process (20 minutes). After that, the covariance features were calculated in 15 minutes, the model trained in 5 minutes and the prediction done in less than a minute.

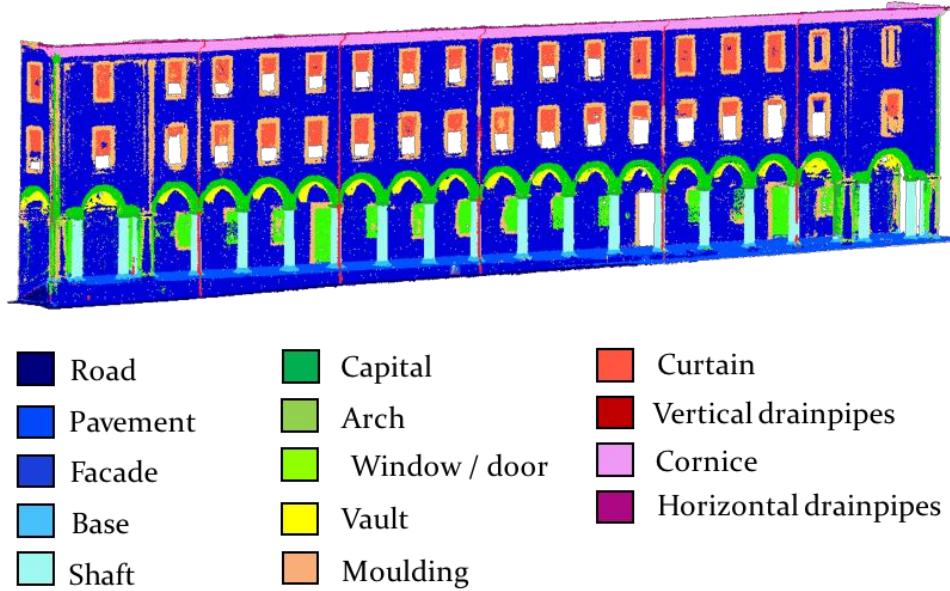


Figure 7.22. Classification results achieved training a Random Forest classifier.

An illustrative comparison between a portion of dataset manually annotated (Figure 7.23a), automatically classified with CGAL (Figure 7.23b) and with ad hoc features (Figure 7.23c) is shown below.

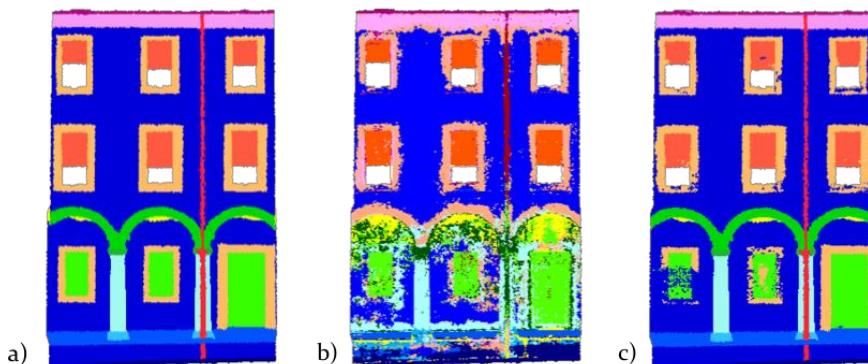


Figure 7.23. A portion of the Bologna dataset: manually annotated (a), classified with CGAL (b), classified with ad hoc features (c).

The development of a new methodology designed for architectural case studies was proved to be helpful for the classification purposes. All the different semantic categories could have been identified and separated (Figure 7.24), even if with certain ones there are still some problems to solve. Indeed, such a result is an excellent point to start, instead of manually segmenting the million-point data.

Part 3. Case studies

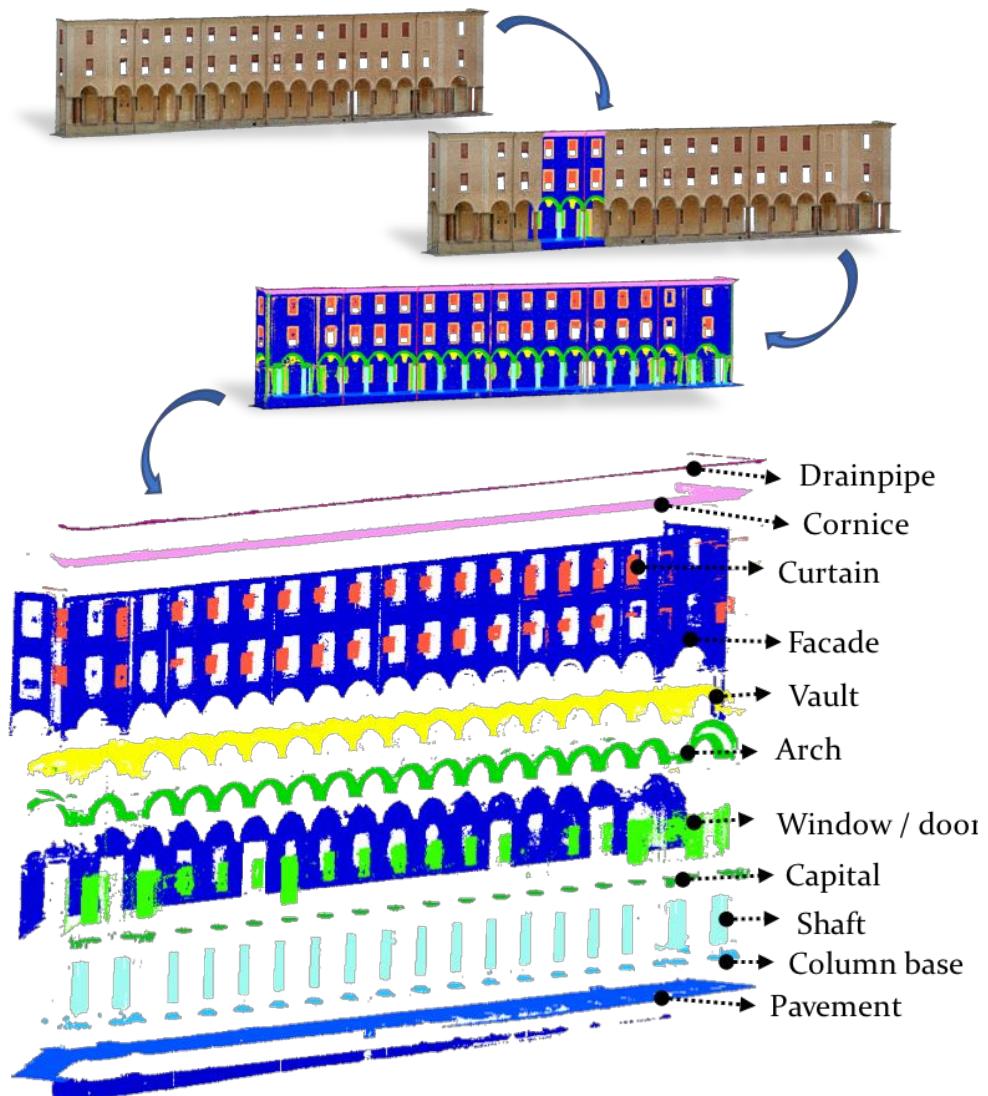


Figure 7.24. Exploded view of the Bologna dataset after the automated classification.

Chapter 7. Geometry-based classification results

REFERENCES

- ETH Zurich Random Forest Template, 2015. [Online] Available at: www.prs.igp.ethz.ch/research/Source_code_and_datasets.html [Accessed May 15th 2019].
- Fiorillo, F., Fernández-Palacios, B. J., Remondino, F., and Barba, S., 2013. 3D Surveying and modelling of the Archaeological Area of Paestum, Italy. *Virtual Archaeology Review*, Vol. 4(8), pp. 55-60.
- Giraudot, S. and Lafarge, F., 2019. Classification. In *CGAL User and Reference Manual*. CGAL Editorial Board, 4.14 edition.
- Özdemir, E., Remondino, F., 2019. Aerial point cloud classification with deep learning and machine learning algorithms. *Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Science*. Proceedings Geospatial Conference 2019, in press.
- Remondino, F., Gaiani, M., Apollonio, F., Ballabeni, A., Ballabeni, M. and Morabito, D., 2016. 3D documentation of 40 kilometres of historical porticoes - the challenge. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 41.

CONCLUSION AND FUTURE WORKS

In the first part of the thesis, a review about 2D and 3D segmentation and classification algorithms was presented. Although not fully exhaustive, this study reported the most popular approaches suitable for both the geospatial and the heritage community. The methods were divided into two main categories, traditional and machine learning approaches, according to their core line. The key difference among these approaches is the method or criterion used to measure the similarity between the points and hence make the grouping decisions.

This deep literature review revealed an evident lack of classification approaches designed for the architectural environment, contrarily to the high level reached in the geospatial field. Therefore, the development of a process able to automatically classify archaeological/architectural 3D data was considered to be highly demanded.

To bridge this gap, two different classification methods based on a supervised machine learning pipeline have been developed and tested. Both methods allow to semantically enrich 3D heritage data, either working on the textures of the models (texture-based approach) or the geometry of the point clouds (geometry-based approach).

With the texture-based approach (described in Chapter 4 and explored in Chapter 6) experts such as archaeologists, restorers, etc., can automatically classify the textures of the heritage objects, starting from small annotations of the classes. The results of the 2D classification can then be projected and visualised onto the 3D geometries for a better understanding.

The geometry-based approach (described in Chapter 5 and explored in Chapter 7) provided a general and straightforward method to classify heritage point clouds, through the use of ad hoc features, i.e. the covariance ones, able to highlight the architectural elements. The choice of the optimal neighbourhood radius for the covariance features extraction is based on the

Conclusions and future works

knowledge of a few essential measures of the object analysed. Indeed, neighbourhood sizes are derived from simple proportional and dimensional rules, typically used for the construction of classical architectures and re-proposed in the different heritage buildings of the following centuries. The experiments demonstrated that it's not necessary using many features extracted at different scales, to get the most accurate classification results (as standard practice in the geospatial field). As future works, further tests will be carried out to (i) verify the applicability of the developed methodology to more complex and not repetitive structures, as well as (ii) identify and test further geometric rules for the classification of buildings with different architectural styles.

For both the aforementioned approaches, the existence of a wide variety of building techniques and the usage of diverse ornamental elements were found to act as obstacles in generalising the classification techniques to heritage case studies. Besides, monuments can present different types of decays, depending on their exposure under various conditions, hence increasing the difficulty of the classification tasks.

To mitigate these issues, the presented machine learning-based approaches were proved to be beneficial for classifying large, varied and complex scenarios, provided that the training datasets are sufficiently large and well-assorted among the classes.

With regards to the choice of the most appropriate approach, the author believes that in the heritage field each case study must be treated individually. This means that for every object under investigation, it is essential understanding which are the classification purposes, its needs, and the required classes. Once those are clarified, the most fit-for-purpose approach can be applied (texture- vs geometry-based), depending on the object's shape, complexity, dimension, presence of good texture, etc.

The main strengths of the proposed methods can be summarized as follows:

- reduced manual input;
- short time to classify big data;
- high levels of accuracy;
- automatic detection of small cracks or details that can hardly be caught by the human eye;
- possibility to map the decays on UV maps and then re-project them onto the 3D models;
- possibility to compute the areas that each class occupies, thus deriving useful data for monitoring and restoration purposes;
- automatic recognition of similar architectural elements in vast datasets, that can be potentially linked to parametric families within HBIM environments;

Conclusions and future works

- applicability of the approaches to different kind of buildings, monuments or any other type of 3D data.

From the analysed case studies, the following remarks and open issues can be finally drawn:

- feature extraction procedures require high computational costs. To accelerate the computation, in most cases it was necessary to subsample the datasets. Therefore, to keep the original point cloud density, an approach based on different levels of details will be developed in the future. The idea is to first discriminate the macro categories on a subsampled point cloud (i.e., floor, facade, roof). Then each category will be iteratively divided into subclasses (i.e., façade → windows → window frames + glasses), processing step by step more dense clouds;
- the higher the radius of extraction and the number of the features are, the higher is the time necessary for the computation. This problem was partially solved, introducing the selection of a few ad hoc features, that speed up the processing time;
- within the heritage field, the training sets can hardly be transferred to other datasets, due to the high variability of the input data. To provide an element of solution, for future works the author aims to work across different heritage buildings and generalize the geometry-based classification approach w.r.t some basic classes (e.g. windows, doors, and columns). In order to accomplish this task, it will be necessary to increase the number of labelled data. In this regard, the creation of a new benchmark project, that collects annotated architectural data and measure the performance of state-of-the-art algorithms through shared datasets and platforms, will be pursued in the future.

Acknowledgments

ACKNOWLEDGEMENTS

I first gratefully acknowledge the funding received towards my PhD from the Fondazione Bruno Kessler (FBK) PhD fellowship.

I would like to say a very big thank you to my supervisor Dr Fabio Remondino for the great opportunity he gave me to work in his research group. Thank you for all the support and encouragement you gave me during these three years. Without your guidance and constant feedback, this PhD would not have been achievable.

I'm sincerely grateful to Dr Prof. Luca Cipriani and Dr Prof. Filippo Fantini, who first introduced me to the world of photogrammetry, and then convinced me, three years ago, to apply for the PhD position. If it weren't for you, I wouldn't be where I am now.

I sincerely thank all the fellow lab mates I had the opportunity to meet during the last three years at the 3DOM research unity. Thanks for the stimulating discussions, the constant support, and for all the fun we have had together.

To conclude, I cannot forget to thank my family for all the unconditional support in these full years. This thesis is dedicated to you.

Data source acknowledgements

I would like to acknowledge the Accademia Adrianea and the University of Bologna (Dept. Architecture) for providing the datasets of Pecile's wall in Villa Adriana and Bologna's porticoes.

I'm thankful to Soprintendenza per i Beni Archeologici dell' Etruria Meridionale and the Etruscan National of Villa Giulia in Rome for the possibility to work on the Sarcophagus and Bartoccini's tomb case studies.

Finally, a special thanks goes to Fausta Fiorillo (Politecnico of Milano, Italy) for providing the dataset of the Basilica and Temple of Neptune in Paestum.