State-of-the-Art Survey

LNCS 12072

Marinos Ioannides
Giovanni Issini
Daniel Oliveira (Eds.)

# 3D Research Challenges in Cultural Heritage IV

## Risk Prevention and Monitoring Methods

OPEN ACCESS

Springer

The series Lecture Notes in Computer Science (LNCS), including its subseries Lecture Notes in Artificial Intelligence (LNAI) and Lecture Notes in Bioinformatics (LNBI), has established itself as a medium for the publication of new developments in computer science and information technology research, teaching, and education.

LNCS enjoys close cooperation with the computer science R & D community, the series counts many renowned academics among its volume editors and paper authors, and collaborates with prestigious societies. Its mission is to serve this international community by providing an invaluable service, mainly focused on the publication of conference and workshop proceedings and postproceedings. LNCS commenced publication in 1973.

Marinos Ioannides · Giovanni Issini ·
Daniel Oliveira
Editors

# 3D Research Challenges in Cultural Heritage IV

## Risk Prevention and Monitoring Methods

Springer

*Editors*
Marinos Ioannides
Cyprus University of Technology
Limassol, Cyprus

Giovanni Issini
SABAP AP-FM-MC
Ascoli Piceno, Italy

Daniel Oliveira
University of Minho
Guimarães, Portugal

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-031-93752-1 ISBN 978-3-031-93753-8 (eBook)
https://doi.org/10.1007/978-3-031-93753-8



The cover image depicts a real-virtual model of the Castel di Luco complex iun Acquasanta Terme, Italy, and was created by the working group of the EU OPHERA project grant number 607601-CREA-1-2019-1-IT-CULT-COOP1, under the European Union's Creative Europe Program. Ownership: EU OPHERA project – UNIFE.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Preface

In an increasingly fragile world, cultural heritage faces constant threats – from catastrophic earthquakes and environmental disasters to armed conflicts and the accelerating impact of climate change. Across the Middle East, Africa, Eastern Europe, and other vulnerable regions, historic buildings, monuments, and artifacts are disappearing at alarming rates. In response, the convergence of **3D digitization, structural analysis**, and **participatory technologies** is redefining how we protect, restore, and reimagine our shared heritage.

This volume issue brings together leading-edge research and multidisciplinary practices demonstrating how **digital tools** – from 3D scanning and photogrammetry to HBIM (Heritage Building Information Modeling), Digital Twins, and Extended Reality (XR) – are transforming the way we understand and care for tangible heritage. These technologies not only record the physical form and material conditions of sites but also enable simulations of deterioration, design interventions, and reconstruct with accuracy and empathy.

Crucially, **structural analysis integrated within 3D models** offers predictive insights into the stability and longevity of heritage sites. By analyzing stress factors, material decay, and environmental risks, especially in earthquake-prone areas, we can move from reactive conservation to **preventive preservation**– mitigating potential damage and informing smarter, safer restoration practices.

A central theme of this issue is the recognition of post-disaster restoration – not merely as a technical operation but as a **living cultural process**. In the aftermath of earthquakes, restoration and reconstruction efforts often provide opportunities for local communities to **recover, reinterpret, and renew their cultural identity**. The act of rebuilding fosters a collective narrative of resilience and pride. To support this process, there is a growing need to define tools and methodologies that enable **public access and cultural use of damaged sites during their restoration**, allowing them to remain part of everyday life even amid repair. Making the **scientific and technical content of restoration accessible** to broader audiences through innovative visualizations, interactive platforms, and storytelling is essential. Digital tools should not only serve experts but also empower the public, allowing for **participation both on-site and remotely via the web**. Furthermore, the integration and publication of restoration data in accessible formats can serve multiple audience groups – including local communities, researchers, policymakers, and educators. This participatory model also calls for the **training of cultural operators and restoration professionals** in effective communication and engagement methods, bridging the gap between expert knowledge and public understanding. Simultaneously, prioritizing **education for younger generations** is critical, fostering awareness of best practices in architectural conservation and inspiring future stewardship of cultural heritage.

In an era where digital and physical realities increasingly intertwine, 3D digitization is no longer a technical luxury - it is a **cultural imperative**. This volume advocates a holistic approach to heritage management, combining cutting-edge technology with local knowledge, risk analysis with creative reuse, and positioning the act of restoration as a bridge connecting the past, present, and future.

Marinos Ioannides
Giovanni Issini
Daniel Oliveira

# Acknowledgement and Disclaimer

This open access publication was partially supported by the European Commission, grant number 607601-CREA-1-2019-1-IT-CULT-COOP1 - Project OPHERA - Opening cultural heritage to communities during the central-Italy post-earthquake long-term restoration process: digital technologies and new competencies for cultural professionals and by the UNESCO Chair on Digital Cultural Heritage at the Cyprus University of Technology.

However, the content of this publication reflects only the authors' views and the European Commission, the Cyprus University of Technology, Limassol, Cyprus, the Ministry of Culture in Italy and the University of Minho, Portugal are not liable for any use that may be made of the information contained herein.

# Contents

# AI-Driven Analysis in Point Clouds for Archaeological Documentation

Jesús Muñoz-Cádiz[1(✉)], Ramona Quattrini[1], and Rafael Martín-Talaverano[2]

[1] Polytechnic University of Marche, 60131 Ancona, AN, Italy
j.munoz@pm.univpm.it, r.quattrini@univpm.it
[2] University of the Basque Country, 20018 Donostia/San Sebastian, SS, Spain
rafael.martin@ehu.eus

**Abstract.** This paper presents an approach for documenting historical phases in architectural heritage by implementing unsupervised Machine Learning (ML). We employ the RANSAC algorithm for architectural segmentation and K-means to analyze the historical sequence in point clouds using geometric features. Finally, the Extended Matrix (EM) tool within a native IFC environment is used for the archaeological metadata linkage standardization. We have tested our approach using several constructive elements of the San Isidoro complex, in León (Spain).

**Keywords:** unsupervised machine learning · geometric features · stratigraphy · clustering · K-means

## 1 Introduction

Unstructured point clouds obtained via static terrestrial laser scanning consist of a huge amount of data, often underexploited. Processing and managing these three-dimensional (3D) objects are complex, particularly in architectural and archaeological heritage. The historical analysis depends strictly on in-depth methodologies such as Archaeology of Architecture (AA), where stratigraphical analysis is commonly presented through orthoimages and computer-aided design drawings. In such cases, the point cloud is just used for visualization and measurements despite the time-consuming effort required for their generation.

Although many improvements have been made in recent years, integrating archaeological analysis into the Heritage Building Information Modeling (HBIM) workflow remains challenging. Artificial Intelligence (AI), specifically Machine Learning (ML), has been adopted in Cultural Heritage (CH) through the implementation of supervised methods i.e. classification and segmentation of building elements. While computationally less intensive than supervised methods, clustering has not yet been explored for archaeological analysis in point cloud data. This kind of unsupervised ML facilitates the identification of meaningful patterns without labeled data. Clustering is defined [1] as a kind of unsupervised ML that implements the basis of partitioning methods based on similarity and dissimilarity functions. The K-means algorithm proposed by S. Lloyd [2]


M. Ioannides et al. (Eds.): *3D Research Challenges in Cultural Heritage IV*, LNCS 13577, pp. 83–96, 2025.
https://doi.org/10.1007/978-3-031-93753-8_6

is an example of this approach, specifically a centroid-based clustering that depends on a given number of clusters in the initialization step.

AA is based on stratigraphical analysis to identify the historical and constructive phases of the building. The fundamentals of stratigraphy were established in [3] and then adapted to the implementation of built heritage [4]. In this sense, the information linked to 3D objects is essential for CH documentation and influences conservation guidelines. However, integrating archaeological data in the Industry Foundation Classes (IFC) schema poses specific limitations due to the lack of HBIM standards.

### 1.1 Problem Definition and Contribution

On-site observations and 2D representations comprise archaeological analysis and documentation work. While integrating this data into a 3D environment is time-consuming, the specific conditions of archaeological analysis pose additional challenges. The specific features of historical buildings, i.e. the lack of standardization of constructive elements, the reuse of materials, or the transformations that take place throughout the whole life of the building often hinder interpretation and the automation of tasks. Consequently, point cloud segmentation in the HBIM approach remains a tedious process.

Given the gap of unsupervised ML applications in archaeological documentation, we propose an adapted approach to point cloud analysis. Unlike existing approaches primarily focused on architectural or damage assessment, our approach aims to serve as an initial step that could distinguish historical phases of buildings based on their geometric features. To address this, we first employ the RANSAC algorithm for the architectural segmentation of complex and irregular geometries. Once the architectural elements are segmented, we use the K-means algorithm to group each point cloud into archaeological phases based on its geometric characteristics. After that, the correlation matrix enables the comprehension of geometric feature relationships intended to be generalized for other samples. This method is particularly designed to support the archaeological phase recognition mainly in ashlar masonry walls.

Given that extracting meaningful and tangible information is a key aspect of data mining, the contributions of this paper are as follows: i) Using the K-means algorithm for the Archeology of Architecture approach through the analysis of unlabeled point cloud; ii) combining it with the Scan-to-MesHBIM [5] methodology, which enables obtaining highly detailed 3D objects. These applications have been implemented in the church of the Real Collegiate of San Isidoro (Leon, Spain).

## 2 Related Literature

Recent advancements in the 3D archaeological analysis are demonstrated through tools such as Extended Matrix (EM), a graph language representing the Harris Matrix method. It combines graphic representations and metadata associated with 3D objects for analysis and definition of scientific virtual reconstructions [6]. The protocol developed during the last years and tested in many case studies shows an approach strongly related to virtual reconstruction and dissemination.

Regarding the HBIM methodology, many improvements in workflows and tools have been introduced [7]. Authors in [8–10] have tested the potential of documenting archaeological data in authoring BIM environments. The work presented in [11] enable historical metadata integration into the IFC schema to improve the effectiveness of conservation tasks in HBIM projects, while stratigraphic analysis has not been included. Abate et al. [12] introduced the need to combine EM and HBIM for documentation activities, but metadata interoperability and geometrical issues during the integration are not addressed.

However, the literature analysis shows a lack between the archaeological analysis, currently performed in two-dimensional (2D), and its representation in 3D through HBIM environments. Since its scenario faces the better management of point clouds, many works related to geometrical features, also named shape descriptors, have been performed for specific analysis. The author in [13] has performed exhaustive state-of-the-art AI algorithms for automating BIM modeling processes based on point cloud data. Implementations of supervised ML algorithms in HBIM approaches include Random Forest [14] or k-nearest neighbors [15], providing good results in the recognition of architectural elements. Other authors [16] have proposed a classification method related to decay according to a range of wall defect types on point clouds. Their method based on two-dimensional (2D) Continuous Wavelet Transformation (CWT) identifies in the point cloud the constitutive units (stone blocks and mortar regions) and then, using supervised ML, these are classified according to decay type.

Nevertheless, the issues about time-consuming labeling tasks with supervised ML have already been highlighted in [17]. In the field of Unsupervised ML, the work presented in [18] has analyzed the clustering algorithms: strengths and weaknesses, comparison, and potential directions. The K-means algorithm has been reviewed for its variety and spread in many fields [19], including extensions such as U-k-means [20], which finds the number of clusters without knowing a priori the optimal $k$. In [21] the K-means algorithm has been incorporated to improve the accuracy and optimization of point clouds. Modern Clustering algorithms are more developed and powerful. In the work presented by [22], a novel weighted kernel K-means algorithm that integrates aspects of spectral clustering is developed, improving computational efficiency and maintaining accuracy in clustering tasks. While these works have improved the data processing of historical assets, they have not solved the problem of archaeological documentation using point clouds in an HBIM workflow.

## 3 Proposed Method

The method proposed for historical documentation using point clouds is divided into three main steps. First, after preprocessing the PCD, architectural segmentation is performed using the RANSAC algorithm. In the second step, geometric features are selected, calculated, and normalized to prepare the data for subsequent analysis. Finally, K-means is applied to cluster the points to guide the 3D meshing step.

### 3.1 Automated PCD Segmentation Using RANSAC

The Random Sample Consensus (RANSAC) [23] is a model-fitting algorithm. It iteratively selects random subsets of points to model the geometric shape of the plane. It searches for the optimal normal vector $\overrightarrow{n}$, which maximizes the number of inliers within a given distance threshold ($\epsilon$). The process continues as follows:

$$P_i = \arg\max_n \sum\nolimits_{j \in C} \left( \frac{|d(n, p_j)|}{\vec{n}} < \varepsilon \right) \quad (1)$$

where $P_i$ represents the set of inliers corresponding to the plane $i$. This set contains all the points that fit the identified geometric model. Outliers cannot be considered since they may represent additional geometry.

The use of RANSAC in architectural heritage is limited to the presence of complex architectural shapes, but its implementation, as commented before, is well-known in scientific literature. However, HBIM approaches do not include it often since parametric and other kinds of modeling are more used during the 3D meshing step, omitting PCD segmentation.

### 3.2 Point Cloud Feature Engineering

In a point cloud $P$, each point is represented by its Cartesian coordinates x, y, and z, related to an established origin. However, each point can hold other values such as geometric features $p_i = \{f_1, f_2, f_3 \ldots f_n\}$, where $f_i$ defines the feature value of a point in a space. Defining reliable features is crucial for archaeological phase recognition, where redundant or irrelevant information can be filtered. The geometric features are inherited data properties that define distance and similarity measurement [24] and include roughness, planarity, linearity, verticality, omnivariance, eigentropy, and the 1st, 2nd, and 3rd eigenvalues. Eigenvalues are mathematical concepts utilized for analyzing behavior, linear transformation, or matrices, providing insights into the distribution of points along different axes or dimensions.

Separately, RGB values can be considered only if good lighting conditions exist. The local neighborhood radius is analyzed for each 3D point in the point cloud establishing a distance-based criteria, and the covariance matrix is used to extract these geometric features. This filter-based feature selection method is applied using the Euclidean distance, a proximity measure that will be extremely presented during the entire process. If point $p_i$ is given by its coordinates $X_i$, $Y_i$, $Z_i$ and point $p_j$ is given by its coordinates $X_j$, $Y_j$, $Z_j$, the Euclidean distance can be calculated as:

$$d(p_i, p_j) = \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2 + (Z_i - Z_j)^2} \quad (2)$$

**Normalization and Log Transformation.** Data normalization involves transforming raw data features to be on a similar scale, affording more stability to the model and robust calculations. Features that consist of large scales can be more dominant than others during the distance calculation, producing distortional results [24]. Histogram shapes are heterogeneous, requiring different normalization techniques for subsequent analysis. For example, the histograms that are close to bell-shaped distribution may be normalized via scaling to the range, also named the min-max normalization method:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{3}$$

where $x'$ is the scaled value, x is the original feature value, $x_{min}$ is the minimum value in the dataset, and $x_{max}$ is the maximum value in the dataset.

### 3.3 K-means Clustering

After normalizing data, accurate similarity calculation between variables is possible. Quantitative input features are utilized in a K-means clustering algorithm to obtain qualitative values. Hyperparameters are defined to control the behavior of the K-means algorithm, which is considered a crucial step for achieving meaningful results.

**Initialization: Random Centroids Selection and Point Assignment.** When choosing the initialization method, the rest of the hyperparameters that should be defined are the number of clusters, maximum iterations, tolerance, and precomputed distances. In the input data the vector $K = \{K1, \ldots, Kn\}$, the integer parameter $k = n$ is the number of the vector elements. Once the number of clusters is defined, each cluster's centroid $\theta_k$ is randomly determined. After this, points are assigned to each closest $\theta_k$ calculating the Euclidean distance of each point to each centroid. The data space is partitioned into regions as Voronoi cells, closest to a particular centroid.

**Optimization: Cluster Re-assignment.** As the centroids shift, the algorithm reevaluates each point selected. Points are assigned to the closer centroid, creating new clusters. This iterative process continues until the data points no longer switch clusters between rounds (it will be refined until they do not change significantly). This process helps to mitigate any bias introduced by the randomness of initial centroid selection. The distance calculation for point assignment is obtained again through the Euclidean distance, and in this stage, it will determine which cluster centroid is closest to each point.

## 4 Experimental Results

In the following section the RANSAC and K-means algorithms are applied for historical documentation in the Real Collegiate of San Isidoro in León, Spain. This implementation, including preprocessing steps, has been conducted on the Google Colab platform through the NumPy and Open3d [25] Python libraries. The final step is to generate 3D surfaces from the PCD analysis into an HBIM project. The test has been carried out in Bonsai, the former BlenderBIM [26], which allows working with native IFC.

### 4.1 The Archaeological Sequence of the Real Collegiate of San Isidoro (León, Spain)

The Real Collegiate of San Isidoro, in León (Spain), is a complex of buildings that dates back to the 1st half of the 11th century when a pre-Romanesque building was erected over the Roman walls of the city. It comprised a church and, at least, two adjacent spaces on the west side: the Pantheon (lower floor) and the Cámara de Doña Sancha (upper floor). The church was partially demolished to build a bigger Romanesque Basilica over it, completed in the 12th century, and the only remaining part of the previous church is the west end. Our work then focuses on the part where the pre-Romanesque pantheon and church relate to the 12th-century Romanesque Basilica.

This area has been analyzed with the methodology of the AA so that its division into historical phases is already known and it is possible to compare it with the results of our analysis [27]. The rich historical sequence of the building directly affects its geometrical features. These spatial properties are chosen to perform our analysis in two specific areas of the complex, as Fig. 1 shows. Santoni et al. [10] is considered the first HBIM implementation, taking special attention to the historical dataset and serving as the baseline for the presented contribution.

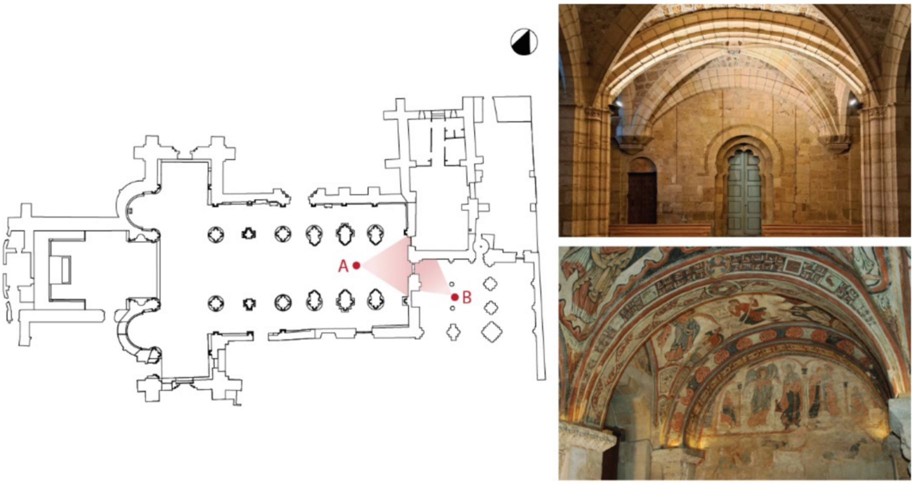


**Fig. 1.** Floor plan of the San Isidoro complex, highlighting the location of the areas (in red) selected for our tests: A) the church's west end bay (top right) and B) the pantheon (bottom right).

### 4.2 RANSAC and K-means for CH Documentation

Preprocessing consists of different tasks to prepare the input point cloud data. This step includes statistical outlier removal and decimation. Then, the normal computational step is performed to ensure it reflects true surface geometry. These aspects are critical for preparing the data for subsequent analysis.

At this point, the RANSAC algorithm is implemented to detect dominant planar features with a segmented portion of the point cloud. Due to the initial stage of the research, only the walls of the selected areas (see Fig. 1) have been considered since they represent the historical sequence of the building better than other architectural elements. A distance threshold was applied to identify inliers for each candidate plane, followed by a filtering stage to discard planes supported by an insufficient number of points. To ensure spatial distinctiveness, a centroid-based comparison was introduced, preventing the detection of redundant or closely overlapping planes.

This method enables the isolation of geometrically significant planar elements, providing a reliable basis for subsequent historical and architectural analysis. Figure 2 depicts the outcome regarding the walls that will be subjected to archaeological analysis. Both cases represent different geometrical conditions, decoration, and states of preservation. For this reason, they are considered good samples for testing K-means capability in various scenarios.

**Fig. 2.** Results of the RANSAC shape detection algorithm applied to the San Isidoro complex: The selected areas have been processed to extract the architectural elements.

The feature engineering step determines which features will be selected, extracted, and used in the training model as input. In our case, RGB values have not been considered for this implementation due to the extreme lighting conditions in indoor religious buildings affecting color gradients. The process begins with the normalization step using the NumPy library. Following feature extraction, the distribution of the nine geometric features was analyzed in both walls. Due to the observed heterogeneity in the shapes, the min-max normalization method was employed in subsequent analysis.

The geometric features mentioned represent some anomalies that correspond with different interventions performed on the wall. Figure 3 represents the normalized histograms of the selected features, surface variation, and verticality, and the point cloud visualization calculated considering a 4 cm neighborhood radius in *Wall_01* and 8 cm for *Wall_03*.

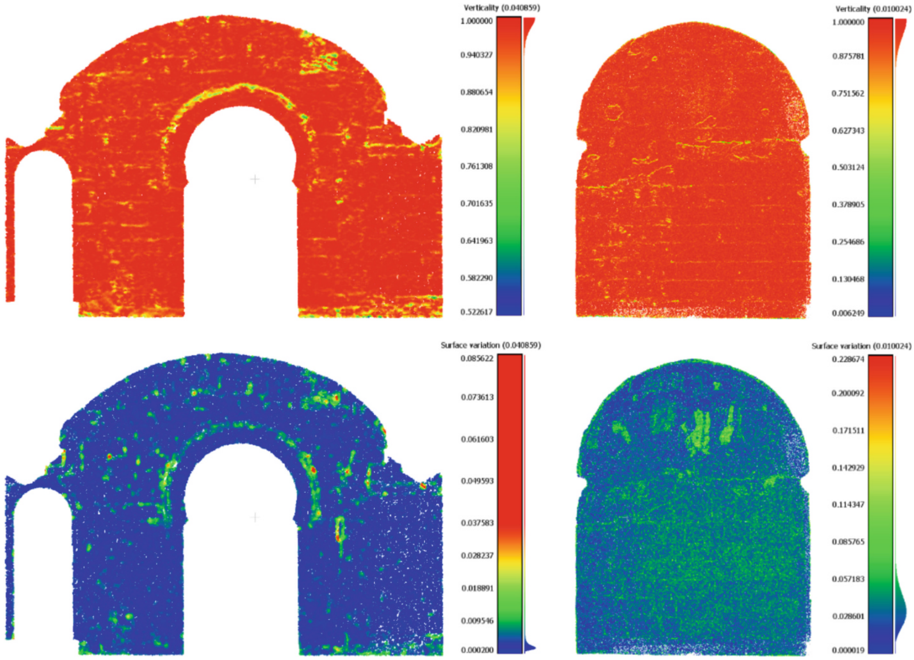


**Fig. 3.** Calculation of verticality (upper part) and surface variation (bottom part) in *Wall_01* and *Wall_03*.

Before applying K-means clustering, the correlation matrix was performed and can be checked in Fig. 6 (Appendix 1). This calculation shows the relationship between the features, in this case, the correlation coefficients offering a standardized measure of the variable's relationship.

**Clustering Results and Validation.** Using the Python library Scikit-learn, K-means clustered in three groups the point cloud of *Wall_01* and in two the *Wall_03*. This generates some inconsistencies according to previous historical analyses [27]. In any case, these clustering results must be evaluated based on various metrics, including 1) cluster cohesion and 2) cluster separation, which refers to how distinct clusters are from each other. To assess cohesion, the Elbow Method was employed by analyzing the inertia, a distance-based metric that measures the Sum of Squared Distances (SSD) between data points and their respective centroid. K-means aims to minimize this value aiming for well-defined and compact clusters.

Figure 7 (Appendix 2) shows the SSD values obtained for varying numbers of $k$ in *Wall_01*. The Elbow method helps to identify the point at which adding more clusters no longer significantly reduces the SSD, thus indicating the optimal number of clusters.

Figure 4 shows the comparison between our analysis and the result of the previous archaeological research. Some areas represent different geometrical feature values that could belong to the historical phases recognized by archaeologists in previous works. Although recognition is not so precise, the result could highlight potential variations in the sequence of transformations of the building.

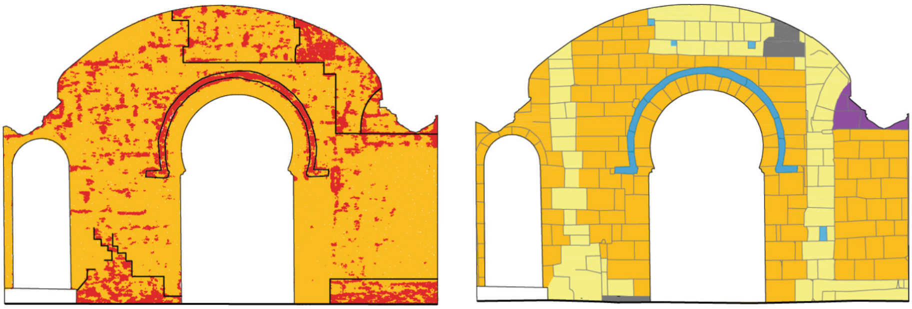

**Fig. 4.** Results of the archaeological phases clustering (red points) applied on *Wall_01* (on the left) and previous archaeological analysis carried out in 2D (on the right).

The historical phase clustering is followed by segmenting the values related to each historical phase, requiring manual corrections since the results are noisy. Then, the 3D meshing step is performed to obtain 3D surfaces with a high level of accuracy. This approach, framed in the Scan-to-MesHBIM methodology, stresses the point cloud and avoids manual modeling.

### 4.3 Integration into the HBIM Project

In a native IFC schema environment, 3D objects represent different subtypes of *IfcBuildingElement* and posterior phases are integrated as *IfcSurfaceFeature*, a solution already implemented in previous works [11] for decay analysis. In this way, an *IfcWall* represents the geometry of the original phase preserved, while the following phases are associated with the object as *IfcSurfaceFeature*.

With our approach, archaeological metadata is manipulated through EM and then integrated into the IFC Schema using the A$^2$Heritage[1] data library capabilities. These data are stored on each 3D object using tailored property sets created to be used in different CH contexts. Each architectural element contains the archaeological phase color as a material, preserving at the same time the RGB values from the point cloud and enabling a realistic scene visualization. The EM tool links geometry to each stratigraphic unit (Matrix Harris) through the yEd Graph editor without affecting IFC data. Figure 5 simplifies the approach of integrating archaeological data into an HBIM framework. Given that the model has preserved the vertex color from the point cloud, it is possible to switch to a realistic view without duplicating building element entities, a commonly presented issue in scientific literature that affects the efficiency of the IFC schema.

[1] The library is accessible for other HBIM users through the following repository: https://github.com/jmc-96/Architectural-Archaeological-Heritage-IFC-Data-Library.

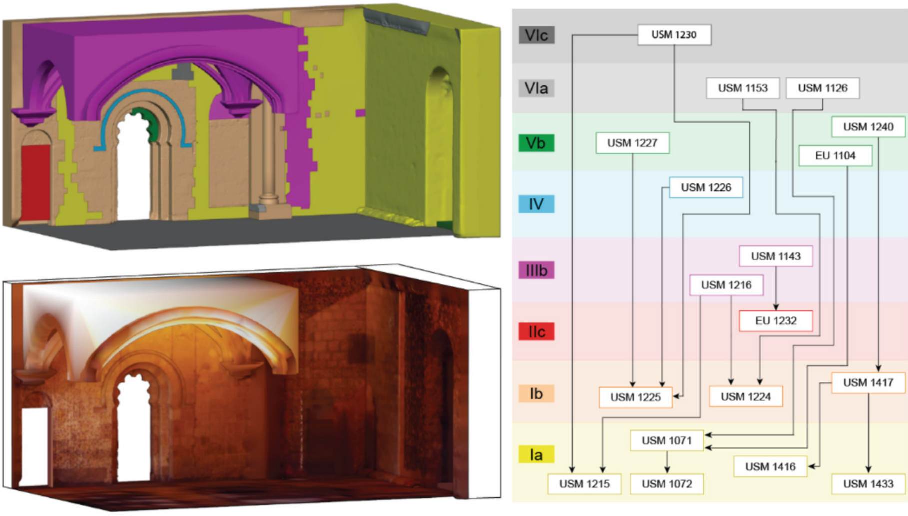


**Fig. 5.** San Isidoro HBIM project: On the left, the archaeological phase view (upper part) and the architectural one (lower part); On the right, the Harris Matrix performed in yEd is linked to 3D objects in the IFC schema.

## 5 Conclusions and Discussion

This study presents a clustering implementation on a point cloud for historical documentation avoiding hand-labelled steps. The paper demonstrates how combining geometric features reflects the degree of change on the point cloud's surface, and, in some areas, it highlights the transformation sequence of the historical building. The method was applied to two walls, a roughly flat element where geometric variations could correspond to different historical phases. In contrast, the clustering is not completely accurate, making necessary some light corrections before the 3D meshing process.

In further experiments, including RGB values should allow us to assess their reliability under favorable lighting conditions. The presented analysis could help the onsite work of archaeologists, so it assesses a tool that could be a useful complement for professionals in some cases. Our aim is not to substitute the on-site analysis of archaeologists but to facilitate and improve it. In this experiment, we tested the algorithm's limitations under these conditions, given that K-means assumes spherical and equal-sized clusters instead of arbitrary shapes. As the problems of funding non-convex clusters may not always hold for archaeological data, other clustering algorithms such as k-medians, k-medoids, or the Gaussian mixture model may be considered in future tests.

The point cloud analysis supports the aims of the Scan-to-MesHBIM approach since it provides the group of points considered as archaeological phases. This scenario allows the implementation of automated 3D meshing steps of each historical surface associated with the "original" wall. With the proposed framework, an HBIM project has been developed and enriched with archaeological data. K-means appears as a partner for the early step of the analysis, guiding the expert to a deeper understanding of the element's historical evolution and then allowing the user to make driven decisions.

**Disclosure of Interests.** The authors have no competing interests to declare relevant to this article's content.

## Appendix 1

Correlation values are commonly filtered between high positive, moderate positive (0.3 to 0.7), low positive (01 to 0.3), non-correlation (–0.1 to 0.1), low negative (–0.1 to –0.1), moderate negative (–0.7 to –0.3) and high negative (–1.0 to –0.7). The correlation between roughness, omnivariance, and verticality is considered as moderate.

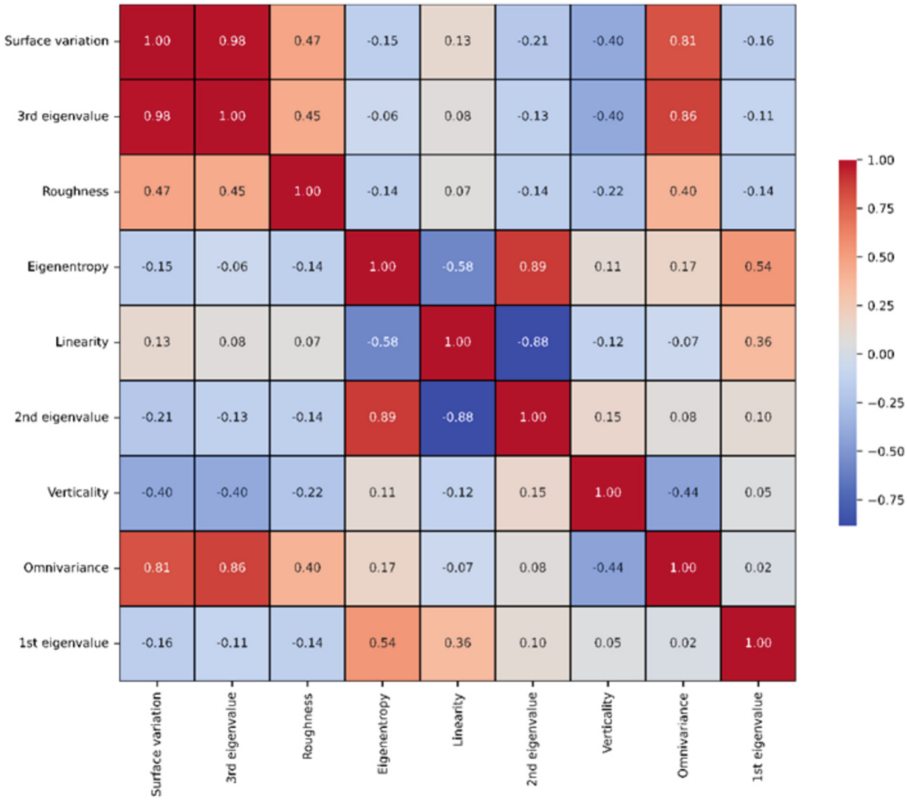


**Fig. 6.** Correlation Matrix of the entire list of geometric features stored in the point cloud of *Wall_01*.

## Appendix 2

Additionally, different cluster results are shown in Fig. 7. From top to bottom, each raw represents different calculations obtained in *Wall_01*: roughness and verticality, roughness and omnivariance, and omnivariance and verticality.

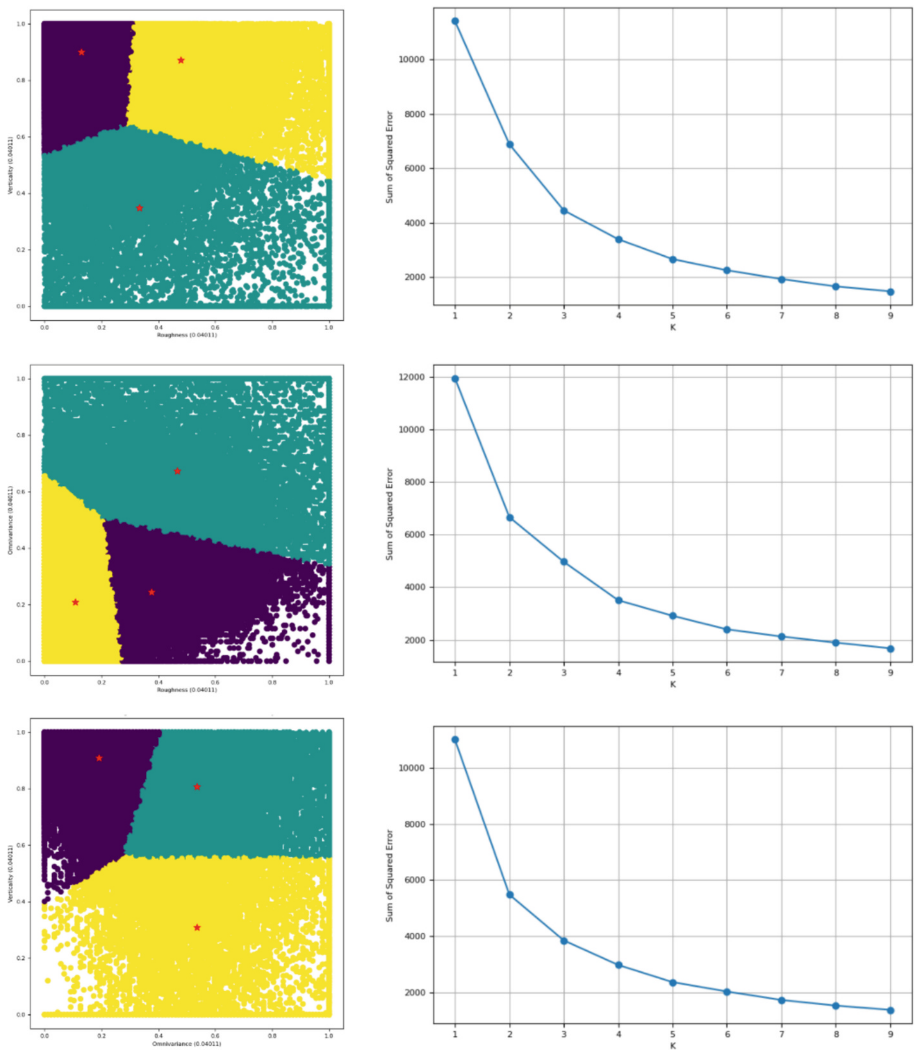


**Fig. 7.** K-means implementation. On the left, the plotted clusters. In the right, the elbow method determines the optimal $k$ number.

## References

1. Jain, A.K., Dubes, R.C.: Algorithms for Clustering Data. Prentice-Hall, Inc. (1988)
2. Lloyd, S.P.: Least Squares Quantization in PCM. IEEE Trans. Inf. Theory **28**, 129–137 (1982). https://doi.org/10.1109/TIT.1982.1056489
3. Harris, E.C.: The laws of archaeological stratigraphy. In: World Archaeology, pp 44–54 (1979)
4. Zoreda, L.C.: Método para el análisis estratigráfico de construcciones históricas o "lectura de paramentos". Inf. Constr. **46**(435), 37–46 (1995). https://doi.org/10.3989/ic.1995.v46.i435.1096

5. Muñoz-Cádiz, J., Quattrini, R., Martín-Talaverano, R.: Scan-to-MesHBIM: implementing knowledge about historical vaulted ceilings with open tools. Rivista Tema 10 (2024). https://doi.org/10.30682/tema100015
6. Demetrescu, E., Ferdani, D.: From field archaeology to virtual reconstruction: a five steps method using the extended matrix. Appl. Sci. (Switzerland) **11**(11), 5206 (2021). https://doi.org/10.3390/app11115206
7. Murphy, M., McGovern, E., Pavia, S.: Historic building information modelling (HBIM). Struct. Surv. **27**(4), 311–327 (2009). https://doi.org/10.1108/02630800910985108
8. Mammoli, R., Mariotti, C., Quattrini, R.: Modeling the fourth dimension of architectural heritage: enabling processes for a sustainable conservation. Sustainability (Switzerland) **13**, 5173 (2021). https://doi.org/10.3390/su13095173
9. Martín Talaverano, R., Murillo Fragero, J.I., Utrero Agudo, M.D.L.Á.: Reflexiones y criterios relativos a la creación de modelos BIM de edificios históricos.Arqueología de la Arquitectura (2021).https://doi.org/10.3989/arq.arqt.2021.005
10. Santoni, A., Martín-Talaverano, R., Quattrini, R., Murillo-Fragero, J.I.: HBIM approach to implement the historical and constructive knowledge. The case of the real Colegiata of San Isidoro (León, Spain). Virtual Archaeol. Rev. **12**(24), 49–65 (2021). https://doi.org/10.4995/VAR.2021.13661
11. Clini, P., Mariotti, C., Angeloni, R., Muñoz-Cádiz, J.: Architectural heritage digital representations for conservation strategies. ISPRS J. Photogrammetry, Remote Sens. Spat. Inf. Sci. 111–118 (2024)
12. Abate, N., Prodomo, M., Frisetti, A., Masini, N.: HBIM and Extended Matrix for the archaeological reconstruction of the excavated medieval contexts: the case of the castle of Monteforte Irpino (AV). In: 2022 IMEKO TC-4 International Conference on Metrology for Archaeology and Cultural Heritage, MetroArchaeo 2022 (2022)
13. Cotella, V.A.: From 3D point clouds to HBIM: application of artificial intelligence in cultural heritage. Autom. Constr. **152** (2023)
14. Croce, V., Caroti, G., De, L.L., et al.: From the semantic point cloud to heritage building information modeling: a semiautomatic approach exploiting machine learning. Remote Sens. (Basel) **13**, 461 (2021). https://doi.org/10.3390/rs13030461
15. Weinmann, M., Jutzi, B., Hinz, S., Mallet, C.: Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. ISPRS J. Photogramm. Remote Sens. **105**, 286–304 (2015). https://doi.org/10.1016/j.isprsjprs.2015.01.016
16. Valero, E., Forster, A., Bosché, F., et al.: High level-of-detail BIM and machine learning for automated masonry wall defect surveying. In: ISARC 2018 - 35th International Symposium on Automation and Robotics in Construction and International AEC/FM Hackathon: The Future of Building Things (2018)
17. Zeng, C., Wang, W., Nguyen, A., et al.: Self-supervised learning for point cloud data: a survey. Expert Syst. Appl. **237** (2024)
18. Sarker, I.H.: machine learning: algorithms, real-world applications and research directions. SN Comput. Sci. **2** (2021)
19. Ikotun, A.M., Ezugwu, A.E., Abualigah, L., et al.: K-means clustering algorithms: a comprehensive review, variants analysis, and advances in the era of big data. Inf. Sci. (N Y) **622**, 178–210 (2023). https://doi.org/10.1016/j.ins.2022.11.139
20. Sinaga, K.P., Yang, M.S.: Unsupervised K-means clustering algorithm. IEEE Access **8**, 80716–80727 (2020). https://doi.org/10.1109/ACCESS.2020.2988796
21. Gujski, L.M., Di Filippo, A., Limongiello, M.: Machine learning clustering for point clouds optimisation via feature analysis in cultural heritage. Int. Arch. Photogrammetry, Remote Sens. Spat. Inf. Sci. **46**, 245–251 (2022)

22. Dhillon, I.S., Guan, Y., Kulis, B.: Kernel k-means, spectral clustering and normalized cuts. In: KDD-2004 - Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2004)
23. Schnabel, R., Wahl, R., Klein, R.: Efficient RANSAC for point-cloud shape detection. Comput. Graph. Forum **26**, 214–226 (2007). https://doi.org/10.1111/j.1467-8659.2007.01016.x
24. Iguyon, I., Elisseeff, A.: An introduction to variable and feature selection. J. Mach. Learn. Res. **3** (2003)
25. Zhou, Q-Y., Park, J., Koltun, V.: Open3D: A Modern Library for 3D Data Processing. arXiv:180109847 (2018). https://doi.org/10.48550/arXiv.1801.09847
26. IfcOpenShell Project (2019) Bonsai. https://bonsaibim.org/. Accessed 7 May 2024
27. Utrero-Agudo, M., Murillo-Fragero, J.I.: San Isidoro de León. Construcción y reconstrucción de una basílica románica. Arqueologia de la Arquitectura **11** (2014). https://doi.org/10.3989/arq.arqt.2014.011