



HAL
open science

Clustering for the Analysis and Enrichment of Corpus of Images for the Spatio-temporal Monitoring of Restoration Sites

Laura Willot, Dan Vodislav, Valerie Gouet-Brunet, Livio De Luca, Adeline Manuel

► **To cite this version:**

Laura Willot, Dan Vodislav, Valerie Gouet-Brunet, Livio De Luca, Adeline Manuel. Clustering for the Analysis and Enrichment of Corpus of Images for the Spatio-temporal Monitoring of Restoration Sites. 5th Workshop on AnalySis, Understanding and ProMotion of HeritAge Contents (SUMAC '23), ACM Multimedia 2023, Nov 2023, Ottawa, Canada. pp.39-47, 10.1145/3607542.3617353 . hal-04283568

HAL Id: hal-04283568

<https://hal.science/hal-04283568>

Submitted on 24 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Clustering for the Analysis and Enrichment of Corpus of Images for the Spatio-temporal Monitoring of Restoration Sites

Laura Willot
ETIS / ENSEA, CNRS, CY Cergy Paris
Université
Cergy-Pontoise, France
LASTIG / Univ. Gustave Eiffel, ENSG,
IGN
Saint-Mandé, 94160, France
MAP / CNRS, Ministère de la Culture
Marseille, France
laura.willot@cyu.fr

Dan Vodislav
ETIS / ENSEA, CNRS, CY Cergy Paris
Université
Cergy-Pontoise, France

Valerie Gouet-Brunet
LASTIG / Univ. Gustave Eiffel, ENSG,
IGN
Saint-Mandé, 94160, France

Livio De Luca
MAP / CNRS, Ministère de la Culture
Marseille, France

Adeline Manuel
MAP / CNRS, Ministère de la Culture
Marseille, France

ABSTRACT

After the fire that took place in the cathedral Notre-Dame de Paris, a Scientific research group composed of 9 Working Groups was set up to study the building. With the aim of creating a digital ecosystem for spatio-temporal monitoring of a restoration site, efforts have been put to develop methodologies for data management and enrichment. In line with this approach, this work focuses on the exploration, analysis and enrichment of a corpus of images and annotations of the cathedral, with regards to their temporal, spatial and semantic aspects, in addition to their visual content. Using prevailing clustering methods focusing on different criteria, we obtained several clusters of images from which new information can be inferred. By exploring these complementary aspects, we identified similarity links between the images. Further work will focus on multimodal analyses of the corpus.

CCS CONCEPTS

• **Information systems** → **Clustering and classification**; *Similarity measures*; • **Computing methodologies** → *Matching*.

KEYWORDS

Cultural Heritage, Semantic analysis, Multimodality, Data enrichment

ACM Reference Format:

Laura Willot, Dan Vodislav, Valerie Gouet-Brunet, Livio De Luca, and Adeline Manuel. 2023. Clustering for the Analysis and Enrichment of Corpus

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
SUMAC '23, November 2, 2023, Ottawa, ON, Canada

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0279-2/23/11...\$15.00
<https://doi.org/10.1145/3607542.3617353>

of Images for the Spatio-temporal Monitoring of Restoration Sites. In *Proceedings of the 5th Workshop on the analySis, Understanding and proMotion of heritAge Contents (SUMAC '23), November 2, 2023, Ottawa, ON, Canada*. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3607542.3617353>

1 INTRODUCTION

1.1 Context

On 15 April 2019, a fire started in the Notre-Dame de Paris cathedral, destroying most of its roof when the spire collapsed. At first, generating a deep emotion from people all around the world, it quickly became an unprecedented and unique opportunity for the scientific community to study the building.

The research group for the study of Notre-Dame de Paris has been organized around 9 thematic Working Groups (WGs) with more than 175 researchers from disciplines as various as archaeology, history, architecture, computer science, physics, chemistry, etc. These researchers produce a massive quantity of heterogeneous data which led to the introduction of the n-dame digital ecosystem [7] developed to bring together new tools and methods for the analysis of these data, with regards to their spatial, temporal and morphological characteristics [8, 9].

In this work, we intend to approach the problem of analyzing, structuring and enriching massive corpus of images of Cultural Heritage (CH) restoration sites, with the help of the construction of similarity links relying on different modalities and clustering methods, applied to all three dimensions identified here – temporal, spatial and semantic – and to their visual content.

1.2 Challenges

One of the definitions of Big Data involves three distinctive features: volume, velocity and variety. Although not applicable as such in the humanities (this is particularly true for the velocity property) [25], this definition highlights three characteristics that can be found in the data produced by the researchers working on the study of Notre-Dame de Paris and represent different challenges to overcome.

First, the massive quantity of data (the volume property, relatively to what humanities produce) makes it nearly impossible for a human being to comprehend and analyse all the data available. Automatic methods and tools are needed to allow any user to have a clear picture of what is the data and how each entry interacts with the others. Second, the continuous production of data (even though not as constant and instantaneous as in other domains, this production can be considered as a reduction of the velocity property to its incremental nature) needs to be addressed by the implementation of methods that can be applied to an ever-increasing amount of data. And third, the data produced is particularly heterogeneous (the variety characteristic) and of a 'real world' nature: each sample, description, image, etc. is a human-produced element in the specific context of a restoration site, in our case the restoration of Notre-Dame de Paris. Hence the common digital framework the Digital Data WG is working on [7].

1.3 Objectives

This work focuses on the photographs acquired for the spatio-temporal monitoring of the cathedral's restoration. This specific item has been chosen for its inherent transversal quality, through time and disciplines. Indeed, many pictures have been taken before and after the fire, allowing a comparison between the different states of the cathedral, including the evolution of the restoration work done in the building. Furthermore, images are the carrier for many more information than the plain visual content they bare: scientists use pictures to keep track of samples' extraction sites, to identify the differences and similarities between several elements of a similar nature, across time and space, etc. Hence, images record activities, observations, steps of workflows or documentation. Exploring the potential of this *implied* information is at the core of this study. We want to make this information *explicit* or at least to uncover the potential in terms of analysis and processing for enrichment.

In that respect, images can be enriched through the use of diverse metadata, falling under one of the following dimensions: time, space and semantics. Thus, we need to consider not only the visual content of the images, but all associated metadata that allow us understand more than what the plain visual content of the images. With an appropriate common framework structuring the data, it is possible to develop and integrate automatic methods to ease the analysis and enrich the corpus.

We will shortly introduce work related to the structuring and analysis of massive corpus of heterogeneous CH data in Section 2. Then, we will present in Section 3 the specific features of the data we focus on. Sections 4 and 5 address the methods used to analyse our data – namely, clustering methods – and the results we obtained. Finally, we conclude and present our future work in Section 6.

2 STATE OF THE ART

The documentation of CH objects [12] and processes [6] is a major concern for the CH community, especially as the volume of data increases. A well-documented digital CH item can be easily found and used, which is essential for a good data management. However

the task of manually adding metadata to the assets can be time-consuming, hence the development of automatic processes for CH data enrichment. These methods can be divided into two groups: the learning-based techniques and the 'handcrafted' ones.

Several of the methods imply the use of Machine Learning (ML) or Deep Learning (DL) to enrich datasets, whether supervised (with pre-determined labels) or unsupervised (without those labels). The literature on the subject is abundant. [21] introduces and adapts a multimodal ML approach to predict some chosen properties for digital artefacts. [3] proposes a classification and annotation method based on Convolutional Neural Networks (CNN) for visual features learning subsequently merged with textual metadata. [13] presents as well an application of CNNs for the automatic classification of images of architectural heritage in order to help researchers and historians in the documentation process. Nevertheless, these supervised, learning-based approaches require efforts and a pre-defined list of labels to be applied to the images. This can be proved very inefficient when dealing with data still unexplored. Therefore, data exploration with unsupervised tools is a necessary first step to understand the data.

Clustering, generally considered as a specific case of classification, is an unsupervised task used to attribute unknown labels to the elements. The labeling is done with regards to a certain similarity shared among the elements [24]. This type of task can be of great use to gather some understanding of the data in the early stages of exploration [10] especially as clustering is often expected to produce interpretable results [20]. As highlighted by [20, 24, 28], because of a lack of consensus on the definition of 'cluster', several classifications and taxonomies of clustering methods have been developed and express the variety of techniques that can be used to create clusters. The choice of technique depends on both the type of data available and the expected results. Therefore, a good understanding of the characterization of the data at hand is necessary before turning to an exploration of the relations between these data.

3 DATA: IMAGE CHARACTERISTICS

In this section, we propose to briefly inventory the data and metadata characterizing the images of the corpus, starting with their visual content and going on with the three other dimensions of interest. To introduce these data, we will begin with a description of the data sources for our case study.

3.1 Data Sources

All along their life cycle, from their production to their archiving, images are enriched through diverse methods (for instance, manual tagging to describe an image content with keywords).

In our case study, most of these data come from Aioli, "a reality-based 3D annotation platform, which allows a multidisciplinary community to build semantically-enriched 3D descriptions of heritage artefacts from simple images and spatialized annotations" [1]. In this platform, a user can create a project, usually limited to a single object or an area of interest, from a set of pictures acquired following a photogrammetric protocol. From these pictures, a point cloud is reconstructed and presented in a 2D/3D viewer, along with the spatialized pictures whose position and orientation are now

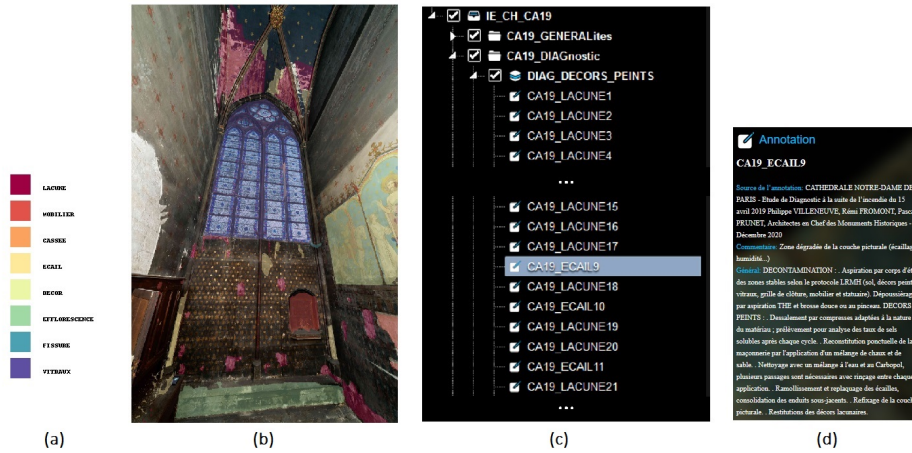


Figure 1: Example of Aioli data

(a) types of regions appearing on the image (b) image from an acquisition process in one of the chapels with the regions in color (c) extract of the hierarchical categorization used to organize the annotations in this project (d) extract of the description sheet associated to the annotation CA19_ECAIL9

known. The user can then annotate any picture by drawing a region of interest on it which is then re-projected on any other image on which this region appears [1].

An Aioli annotation is composed of user-defined elements: the region drawn on the image, a name, and a description sheet. The name belongs to a hierarchy created by the user to organise the different regions of a project in different categories. The description sheet is generated by various contents (such as files, free texts, etc.) chosen by the user to contextualise the regions. Figure 1 presents these elements.

In our case study, the regions of interest and related descriptions (ie, the annotations and their images) come from the digital version of the condition report provided by professionals in the field and were digitized in Aioli [22]. Thus, the annotation tool provides both images and qualitative contextualized information in the form of experts' annotations. The resulting dataset reflects the practice and knowledge in the field of conservation and restoration.

3.2 Image characteristics

3.2.1 *Visual Content.* The visual content of an image is the first accessible data when no metadata is available. It is a low level data that can be automatically analyzed with many approaches coming from image analysis, indexing, computer vision and more generally AI. Note that we can analyze and describe the whole content of the image as well as focus on parts or details.

The visual content is not always enough to describe the images and build a structured corpus of images. Thus, all metadata associated with the images (as presented in the following subsections) is a valuable information that can be used, sometimes in conjunction with the visual content, to deduce new information from the study of the corpus of images.

3.2.2 *Time.* Any image possesses a date of acquisition resulting from the acquisition itself. All images in the corpus is therefore part of a timeline including all known acquisition dates for all these images (Figure 2).



Figure 2: Timeline of the date of acquisition of some of the images

An acquisition date is usually composed of a day and an hour of acquisition, allowing different degrees or levels of analysis: all images taken the same day can be compared to each other, with regards to their hours of acquisition. But the days of acquisition inscribe the images in a larger temporality: one can consider days, weeks or even months to be a more appropriate scale to compare the images to each other. The periods can be identified as specific activities that took place in the cathedral: for instance, a picture of the nave taken during the process of cleaning the space has a date (and time) of acquisition placing this image in the specific day it was taken, but also in the period during which various people came to clear the floor of the nave by recovering the debris that littered it.

3.2.3 *Space.* The spatial dimension may be represented in different ways; in our case, each image can be associated to a 6-dimensional vector (3 values for the position and 3 for the orientation) defined in a reference system common to all the images in the corpus.

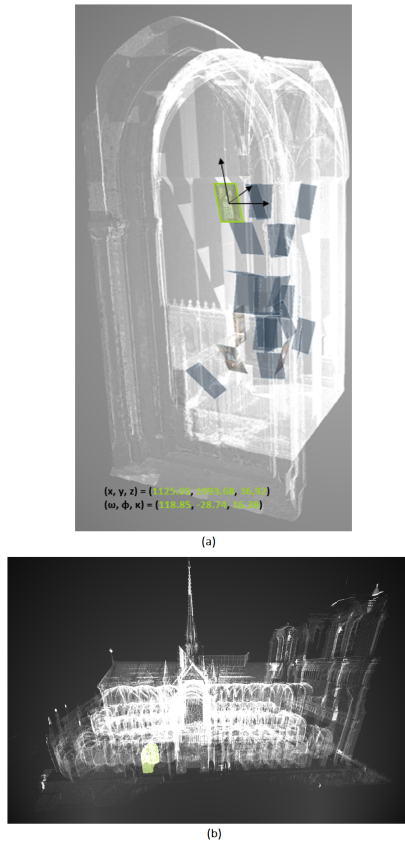


Figure 3: Spatialization of the images

(a) example of an image's coordinates and location in one of the Aioli chapel projects
 (b) location of the same project in the cathedral (in green)

It is possible to consider the image by itself, and thus, get an idea of where it was taken, but it is also possible to look at its position in the corpus, that is, to compare it with the other images' positions. It is also interesting to consider the scale of the restoration project: as the building is particularly huge, it has been "divided" into several spaces (such as the nave, the choir, etc.) which are in turn divided into smaller spaces (Figure 3).

3.2.4 Semantics. Here, the semantic dimension is understood as any string of characters associated with an image describing or enhancing it.

As explained in Section 3.1, we consider the semantic data coming from the Aioli database from which we identified two types of textual data:

- free descriptive texts: user-defined texts from the description sheet associated to a region drawn by the user (Figure 1 (d))
- hierarchical categorization: the tree-like structure of data to which belong the name of the annotation (Figure 1 (c))

Similarly to the temporal and spatial dimensions, different scales of semantic data can be explored. First, descriptive texts are composed of single words giving insights on what's really about the image. But those words can also be part of structures such as controlled vocabularies, glossaries, thesauri and ontologies. These different

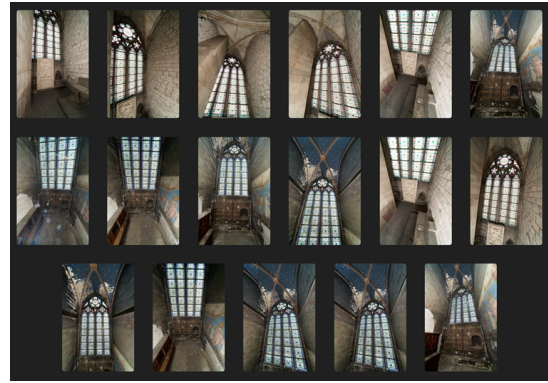


Figure 4: Example of visually alike images from the projects IE_CH_CA19 and IE_NF_CA25

levels of structures give higher and higher degrees of precision and information.

The semi-structured hierarchical categorization of the regions is another interesting data structure to characterize the images and allows different entry point to explore the data related to an image: one could be interested in the leaf node level or any higher level in the hierarchy.

4 METHODS: CLUSTERING

In this section, we present a selection of clustering methods based on the different modalities chosen to explore the corpus of images. These well-proven methods serve as illustrations of image clustering for spatio-temporal monitoring of restoration sites.

4.1 Visual Clusters

The aim of visual clustering is to group images by visual content similarity (Figure 4) to directly create links between the images or to be the basis of further analysis of the other modalities. The data considered is the image itself.

Several algorithms and methods have been developed throughout the years to match images based on their content. Some of these techniques rely on learning (for instance, DELF [16]) whereas others are considered "hand-crafted": SIFT [14], SURF [2] or ORB [23], to name but a few. These image matching algorithms and techniques can be used to create clusters of images sharing similar visual contents.

However, as the learning-based methods might not work as well on our dataset because of the specificity of our domain (CH restoration sites), for this exploratory work we decided to focus on hand-crafted methods: the SIFT and ORB algorithms (Figure 5). Clusters can subsequently be created based on the similarity shown by the matching keypoints between each pair of images.

4.2 Temporal Clusters

Clustering the images by date of acquisition groups images close in time even though they may not share the same visual features.

To achieve this temporal clustering, we consider the day and hour of acquisition of each image. A quick visualization of the data (as in Figure 6) allows us to get an idea of the clusters that can be



Figure 5: Examples of hand-crafted matching algorithms

(left) SIFT (right) ORB

The colored lines correspond to the pairs of matching keypoints

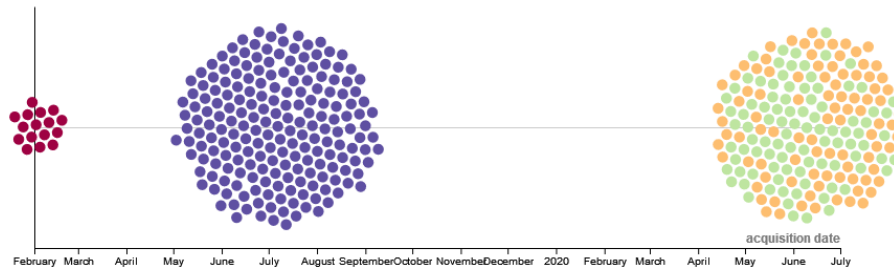


Figure 6: Beeswarm plot of the images' dates of acquisitions (from February 2019 to July 2020)

Each dot corresponds to an image with a known date of acquisition.

The colors correspond to four of the five projects of the dataset (see Section 5.1 for more details on the selected projects).

identified. As a one-dimensional variable, it may not be necessary to use powerful methods such as k-means and related n-dimensional clustering techniques. However, a simple histogram technique is not suitable either as the bins' width is constant whereas we need to be able to create different-sized bins. Thus, we decided to use a visualisation tool for a first exploration of the temporal clusters that could be created with our data. RawGraphs [15] offers several charts for data with a temporal dimension. A beeswarm plot displays items as dots over a horizontal line and uses the vertical space to avoid overlaps. The clusters are then deduced from a visual inspection of the resulting chart.

4.3 Spatial Clusters

The aim of spatial clustering is to group images together by geographical proximity, and for this purpose we have access to the position and orientation of the images.

Based on the division of the cathedral into spaces, we chose to first study the position of the images, regardless of their orientation. Using a Principal Components Analysis as a visualization tool to empirically determine clusters, the data (x, y, z coordinates of the images) is first centered and reduced. Thus, we get a preliminary visualization of the principal components of the position data (Figure 7). As for the temporal clustering, the spatial clusters are determined from a visual inspection of the resulting PCA.

4.4 Semantic Clusters

Semantic clustering aims to highlight the similarities between images that might not be visible directly on the images: in the context

of a restoration site, two images might present different artefacts with the same types of alterations and still be visually dissimilar.

For this exploratory semantic clustering, we focused on the semi-structured hierarchical categorization of the regions drawn on the images, that is any sub-tree of the general tree of alteration patterns presented on Figure 9 (examples of sub-trees are given in Figure 11). Therefore, the resulting clusters take into account only part of the semantic data available in Aioli. But as a complex data structure used to characterize an image, it is a good entry point to study semantic clusters.

To create clusters based on these hierarchical categorizations, a tree-like data structure (a custom Python class) is built for each image. These trees are then compared with a Tree Edit Distance (TED), an adaptation of the string edit distance to ordered labeled trees [26]. The editions considered are: a label modification, a node insertion, and a node deletion. Various algorithms have been developed to compute this TED [4]. We decided to use the Python implementation of the APTED algorithm [17, 18] for its efficiency and ease of use. Each tree is thus compared to all other trees to create a $n * n$ distance matrix, n being the total number of images. Each coefficient of the matrix is the TED computed between the trees (hierarchical categorization) of a pair of images. This matrix is then used to create the clusters with Agglomerative Clustering from the Scikit-learn library [19], a hierarchical clustering method with a "bottom-up" approach [20]. This algorithm takes several parameters in input: a number of clusters and a linkage method (ie: the distance to use between sets of observation). We decided to limit the number of clusters to the number of projects because a smaller number would likely not be enough to differentiate the

images on this criterion, and a higher one would probably create too many groups and hinder the purpose of creating links between the pictures. Following the literature [11], we chose the most used linkage method, namely the average one.

5 EXPERIMENTS & RESULTS

In this Section, we aim to show how the different clustering approaches, used to explore the corpus, can help create an understanding of the data and gather new information.

5.1 Dataset

After three years of work, 240 projects – as defined in the Aïoli platform (cf. Section 3.1) – have been created with more than 9,000 annotations with an average of 49 annotations per annotated project [22]. From these data, we selected 5 representative projects and created a dataset with more than 1000 images and 300 distinct annotations. The statistics for this dataset are presented in the Table 1. This selection of data allowed us to experiment with the clustering analysis methods presented here, while reducing computing times.

5.2 Analysis of the clustering methods

5.2.1 Visual Clusters. The image matching techniques, such as presented in Figure 5, show potential to create visual links between the images. Automating the creation of such links will provide a strong basis to build upon, especially when considering the other dimensions. Indeed, the exploration of visual (dis)similarity can improve the analysis of temporal, spatial, and semantic clusters by providing low-level information on the images grouped together. By taking advantage of the visual links created between the images, an enrichment process can be set up to propagate data from image to image and contribute to an image-level enrichment of pictures with no or few metadata.

On a corpus-level enrichment, preliminary analysis of temporal, spatial, and semantic clusters is necessary to determine where they benefit from these visual links. Hence, the analysis of the other clustering methods that follows.

5.2.2 Temporal Clusters. The visual analysis of the Beeswarm plot presented in Figure 2 shows three main clusters roughly corresponding to the months of February 2019, July 2019 and around June/July 2020. The colors correspond to four different acquisition phases of the cathedral's spaces, two of which took place in a few days in July 2020 (on the 9th and 11th). From what we know (cf. Table 1), we can also see that there were probably two major phases for the acquisitions that took place after the fire: the first one in 2019 for the outside, and the second one in 2020 for the interior spaces. This actually corresponds to a practical reality: right after the fire, the inside of the cathedral was not accessible due to the debris that fell into the nave. Thus, the acquisition process mainly focused on the areas outside the cathedral at that time, before the indoor spaces were cleared and workers were allowed inside.

Based on the dates, we can infer new information: periods of time separated by events such as the start or ending of an acquisition process. We can also consider the different scales of time: one could be interested in the daily acquisition processes whereas another might want to get a more global view of the recording of

the cathedral on a monthly or even yearly scale, for instance. These information can then be used to enrich the corpus of images by characterizing their temporality with higher-level concepts.

As for now, this clustering analysis is just visual and only validated by what people know of the worksite, but it could be improved by automatic methods to precisely define the boundaries of the clusters. Furthermore, these dates cover only the acquisition process and no physical changes that might occur in the cathedral. Thus, it could be interesting to cross this modality with the spatial and visual aspects of the images to determine whether a modification of the spaces happened between the acquisitions of two images. This change might denote that a new cluster boundary is needed and thus create a new level of clusters.

5.2.3 Spatial Clusters. As shown on Figure 7, the PCA provides an illustration of the potential clusters that can be drawn from the analysis of the positions (x, y, z coordinates of the images). We can visually distinguish four main clusters, validated by prior knowledge of the corpus.

Cluster *B* could be further divided into three smaller clusters. These images are part of the project "IE_NF_CB_ExtradosEst" (the extrados of the vaults) and the three sub-clusters actually correspond to the lines followed by the rope access technicians who carried out the acquisitions of this type of difficult-to-access spaces (Figure 8). This is a peculiarity that is not seen in the other projects and differentiates it from these.

Similarly, cluster *C* encompasses both of the exterior projects: *E_NF_TA_TAC_CouverturesNord* (the roof covering between the aisles 29 and 39) and *E_NF_T31_ContrefortF31* (the buttress F31 along the aisle 31). The pictures of these two projects were actually taken in the same zone of the cathedral: the buttress F31 is included in the space between the aisles 29 to 39. This is an illustration of the complexity of the structuring of the spaces of the cathedral and how they are related to each other. In this example, the PCA highlights the closeness between several images which are indeed located in the same range.

Thus, from the images' positions, we can deduce new information in the form of clusters that can be structured into several layers of hierarchy: levels of spaces included in one another. Those spaces can then be used to characterize the images by providing higher-level information than their spatial coordinates.

5.2.4 Semantic Clusters. For the purpose of this demonstration we selected a few annotations from the condition report and aligned the expressions used in this report with a few terms defined in glossaries of stone alteration phenomena and sources [5, 27] (Figure 9). The statistics for this sample of the whole dataset is presented in the Table 2.

As explained in Section 4.4, we only considered the hierarchical categorisation of the regions drawn on the images. This gave us some preliminary results to build upon (Figure 10). The images of a given project mostly belong to several clusters, indicating a variety of alteration patterns in a given space of the cathedral. Likewise, the clusters are mainly distributed over several projects: even though pictures are located in different parts of the cathedral (inside or outside the cathedral, in the nave or in the choir, at the roof level, etc.) or are visually dissimilar, they do share some alteration patterns (Figure 11).

Table 1: Description of the dataset used for the experiments

project name	acq day	I/E	broad loc	typology	#img	#acq time	#spatial	#regions	#annot im
IE_CH_CA19	11/06/20	I	choir	chapel	93	93	93	92	93
IE_NF_CA25	09/06/20	I	nave	chapel	94	94	94	106	94
E_NF_T31_ContrefortF31	03/02/19	E	nave	buttress	16	16	16	31	16
IE_NF_CB_ExtradosEst	--	I	nave	roof	443	0	443	84	433
E_NF_TA_TAC_CouverturesNord	06/07/19	E	nave	roof terrace	596	200	596	18	593
					1242	403	1242	331	1229

The first two columns indicate the name of the project and the day of acquisition (dd/mm/yy) of the images constituting the projects; the third column states whether the images were taken inside (I) or outside (E) the cathedral; the following two columns specify the location and type of space/element; then, a few statistics are given (total number of images, number of images with known acquisition time, number of images with known location, number of regions and number of images with at least one region) per project.

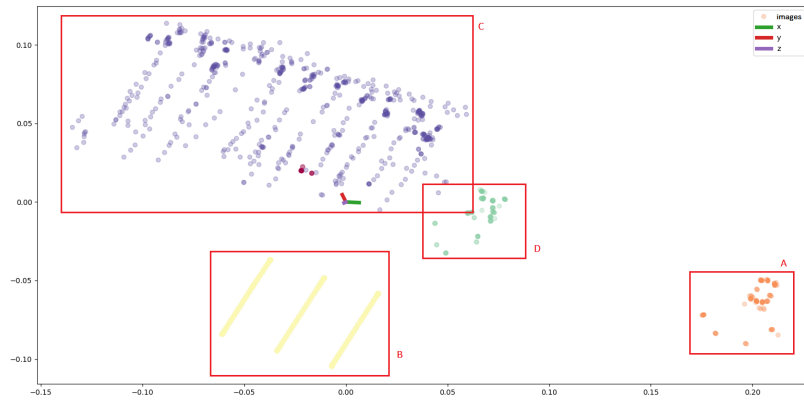


Figure 7: PCA on the x, y, z coordinates of the images
The colors correspond to the five projects for which the positions of the images are known

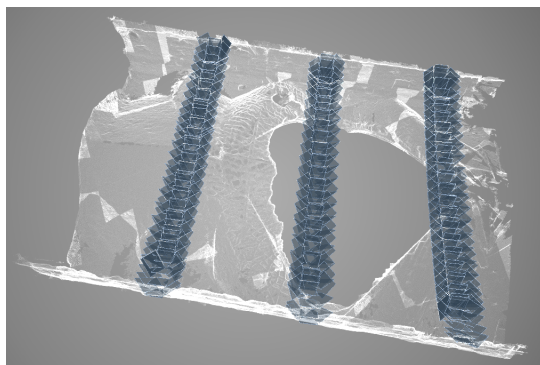


Figure 8: Screenshot of the point cloud and spatialized pictures of the Extrados of the Vaults

Table 2: Statistics for the selected regions of interest

	#regions	#annot img
total	331	1229
selection	166	996

In order to use the Agglomerative Clustering function, we had to set the number of clusters. This is not an easy choice, especially when the aim of this clustering is also to explore the data. One

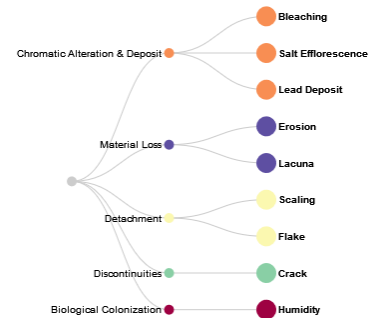


Figure 9: Stone alteration patterns

This is the global tree-like structure of the terms considered. Thus, hierarchical categorization trees of any image will be a sub-tree of this one

potential solution to this could be to take advantage of the "bottom-up" nature of this clustering algorithm and compute the whole dendrogram of clusters instead of stopping it when the given number of clusters is reached. This could be helpful for the analysis of the corpus and create further hierarchical knowledge about the images.

Another improvement to this exploration lies in the use of more semantic data. Taking into account the number and size of each type of regions per image to build the trees and compute the distance

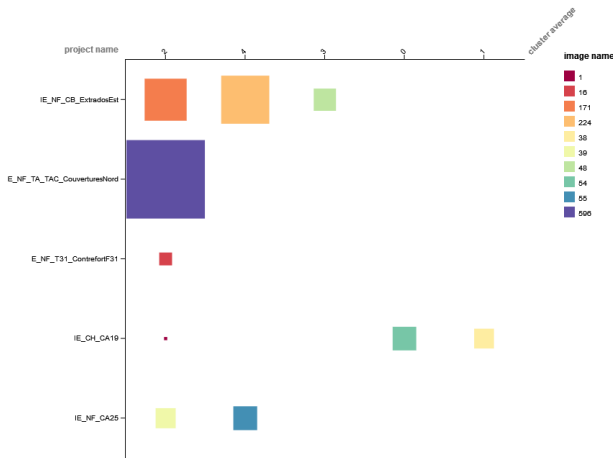


Figure 10: Clusters of images based on hierarchical categorization

The numbers of images are represented per project (lines), per cluster (columns)

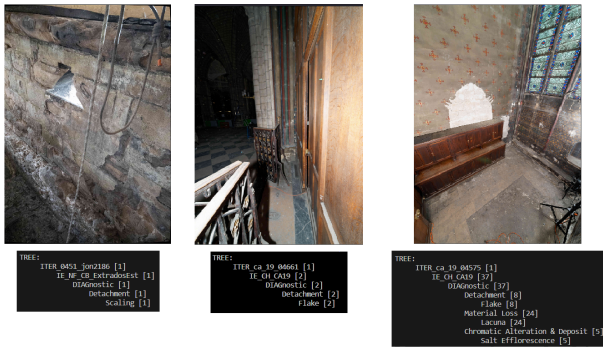


Figure 11: Example of images and their associated annotation-based tree

(left) image from project *IE_NF_CB_ExtradosEst* in cluster 2 (center) image from project *IE_CH_CA19* in cluster 2 (right) image from project *IE_CH_CA19* in cluster 0. The images from the same project are spatially close but the (left) and (center) ones share more similar trees (both have only "detachment" alteration patterns whereas (right) show more of them).

between two of them might help refine this distance. Indeed, two images might have similar trees (ie, the same types of regions appear on both pictures) but one of the region type could be more numerous or bigger on one of the images, suggesting that the alteration is more present on this image than on the other one.

5.3 Global analysis

Due to time constraints, only prior knowledge has been used to evaluate the relevance of the visual and hierarchical clustering of the data. Nevertheless, these different methods of clustering show interesting results for each of the dimensions. They allow to discover new information that can be used to enrich the corpus of images, and can further be utilized to create and answer more

complex queries. These exploratory tools are complementary and serve the same purpose, namely, the analysis and enrichment of the corpus. Such techniques highlight similarities between images that would not have been shown if the visual aspect had been considered alone.

From different types of values characterizing an image (dates, positions, annotations, etc.), it has been possible to create groups of similar images (sharing the same features) and thus, observe them along different axes, understand better their complexity and infer new information which, in turn, can be layered to depict the different scales that can be adopted to study the building. This is further noticeable with a hierarchical construction of the clusters: each level of clusters brings its own information on the images grouped together at that level.

On an image level (as opposed to a corpus level where we consider the information related to the whole corpus or groups of images) the visual links can be further taken advantage of, in conjunction with other types of clusters, to propagate data from image to image: when introducing a new image with no or few metadata to the corpus, the visual similarity between this new image and the ones in the corpus, organised in temporal, spatial and semantic clusters, can be used to deduce metadata to enrich the new image.

6 CONCLUSION & FURTHER WORK

This paper explored a corpus of images based on clustering methods. We showed that, through the clustering of different dimensions (temporal, spatial, semantic, and visual), it is possible to observe them along different axes, to understand better their complexity and then to infer new information to enrich the corpus of images. The analyses of the different dimensions were developed separately to evaluate their own potential to enrich the data. The visual analysis of the visual content, the acquisition dates of the images and the Principal Component Analysis (PCA) of the positions, as well as the agglomerative clustering based on a distance matrix of Tree Edit Distance (TED) computed on the annotations have composed this statistical exploration of the dataset. In order to automatize the data enrichment, the multimodal statistical analysis is a mandatory step to test and validate the workflow and results that will further be enriched by a validation method such as the Silhouette Score. Albeit low level data type, the images in the dataset function as a foundational layer: they confirm or infirm the results visually for the user. The next step is to go from multidimensionality (separate modalities) to combined variables (multimodality and fusion).

ACKNOWLEDGMENTS

This work benefits from a State aid managed by the French National Research Agency (Agence Nationale de la Recherche) under the future investment programme integrated into France 2030, with reference ANR-17-EURE-0021 – Paris Seine Graduate School of Humanities Creation Heritage (École Universitaire de Recherche Paris Seine Humanités, Création, Patrimoine) – Foundation for Cultural Heritage Science (Fondation des Sciences du Patrimoine)

All pictures of the cathedral Notre-Dame de Paris presented in this work are credited under the © Chantier Scientifique Notre-Dame de Paris / Ministère de la Culture / CNRS.

REFERENCES

- [1] Violette Abergel, Adeline Manuel, Anthony Pamart, Isabelle Cao, and Livio De Luca. 2023. Aioli: A reality-based 3D annotation cloud platform for the collaborative documentation of cultural heritage artefacts. *Digital Applications in Archaeology and Cultural Heritage* 30 (Sep 2023), e00285. <https://doi.org/10.1016/j.daach.2023.e00285>
- [2] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. 2006. *SURF: Speeded Up Robust Features*. Lecture Notes in Computer Science, Vol. 3951. Springer Berlin Heidelberg, Berlin, Heidelberg, 404–417. https://doi.org/10.1007/11744023_32
- [3] Abdelhak Belhi, Abdelaziz Bouras, and Sebti Fofou. 2018. Leveraging Known Data for Missing Label Prediction in Cultural Heritage Context. *Applied Sciences* 8, 1010 (Oct 2018), 1768. <https://doi.org/10.3390/app8101768>
- [4] Philip Bille. 2005. A survey on tree edit distance and related problems. *Theoretical Computer Science* 337, 1 (Jun 2005), 217–239. <https://doi.org/10.1016/j.tcs.2004.12.030>
- [5] Philippe Bromblet. 2010. Guide “Altérations de la pierre”. https://www.pierres-info.fr/biblio-taille_de_pierre/alteration-pierre-naturelle.pdf
- [6] N. Carboni, G. Bruseker, A. Guillem, D. Bellido Castañeda, C. Coughenour, M. Domajenko, M. de Kramer, M. M. Ramos Calles, E. K. Stathopoulou, and R. Suma. 2016. Data Provenance in Photogrammetry through Documentation Protocols. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* III–5 (Jun 2016), 57–64. <https://doi.org/10.5194/isprs-annals-III-5-57-2016>
- [7] Livio De Luca. 2022. ERC n-Dame_Heritage. <http://www.ndameheritage.map.cnrs.fr/>
- [8] Livio De Luca and Violette Abergel. 2022. Vers des biens communs numériques pour construire des cathédrales de données et de connaissances pluridisciplinaires sur le patrimoine européen. *Culture Et Recherche* 142 (2022), 104–106.
- [9] Livio De Luca, Violette Abergel, Anaïs Guillem, Olivier Malavergne, Adeline Manuel, Ariane Néroulidis, Roxane Roussel, Miled Rousset, and Sarah Tournon. 2022. L'écosystème numérique n-dame pour l'analyse et la mémorisation multidimensionnelle du chantier scientifique Notre-Dame-de-Paris. Lyon, France, 1. <https://hal.science/hal-03826931>
- [10] Richard O Duda, Peter E Hart, and David G Stork. 1996. *Chap 11: Unsupervised Learning and Clustering* (2nd ed. ed.). John Wiley & Sons, Inc.
- [11] Angur Mahmud Jarman. 2020. Hierarchical Cluster Analysis: Comparison of Single linkage, Complete linkage, Average linkage and Centroid Linkage Method. (2020). <https://doi.org/10.13140/RG.2.2.11388.90240>
- [12] Robin Letellier. 2015. *Recording, Documentation and Information Management for the Conservation of Heritage Places: Illustrated Examples* (1 ed.). Routledge. <https://doi.org/10.4324/9781315793917>
- [13] Jose Llamas, Pedro M. Leronés, Roberto Medina, Eduardo Zalama, and Jaime Gómez-García-Bermejo. 2017. Classification of Architectural Heritage Images Using Deep Learning Techniques. *Applied Sciences* 7, 10 (Sep 2017), 992. <https://doi.org/10.3390/app7100992>
- [14] David G. Lowe. 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 60, 2 (Nov 2004), 91–110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>
- [15] Michele Mauri, Tommaso Elli, Giorgio Caviglia, Giorgio Uboldi, and Matteo Azzi. 2017. RAWGraphs: A Visualisation Platform to Create Open Outputs. In *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter (CHIItaly '17)*. Association for Computing Machinery, New York, NY, USA, 1–5. <https://doi.org/10.1145/3125571.3125585>
- [16] Hyeonwoo Noh, Andre Araujo, Jack Sim, Tobias Weyand, and Bohyung Han. 2018. Large-Scale Image Retrieval with Attentive Deep Local Features. *arXiv:1612.06321 [cs]* (Feb 2018). <http://arxiv.org/abs/1612.06321> arXiv: 1612.06321.
- [17] Mateusz Pawlik and Nikolaus Augsten. 2015. Efficient Computation of the Tree Edit Distance. *ACM Transactions on Database Systems* 40, 1 (Mar 2015), 3:1–3:40. <https://doi.org/10.1145/2699485>
- [18] Mateusz Pawlik and Nikolaus Augsten. 2016. Tree edit distance: Robust and memory-efficient. *Information Systems* 56 (Mar 2016), 157–173. <https://doi.org/10.1016/j.is.2015.08.004>
- [19] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Courville, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
- [20] Pradeep Rai and Shubha Singh. 2010. A Survey of Clustering Techniques. *International Journal of Computer Applications* 7, 12 (Oct 2010).
- [21] Luis Rei, Dunja Mladenec, Mareike Dorozynski, Franz Rottensteiner, Thomas Schleider, Raphaël Troncy, Jorge Sebastián Lozano, and Mar Gaitán Salvatella. 2023. Multimodal metadata assignment for cultural heritage artifacts. *Multimedia Systems* 29, 2 (Apr 2023), 847–869. <https://doi.org/10.1007/s00530-022-01025-2>
- [22] Roxane Roussel and Livio De Luca. 2023. An Approach to Build a Complete Digital Report of the Notre Dame Cathedral after the Fire, using the Aioli Platform. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLVIII-M-2–2023 (Jun 2023), 1359–1365. <https://doi.org/10.5194/isprs-archives-XLVIII-M-2-2023-1359-2023>
- [23] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. 2011. ORB: An efficient alternative to SIFT or SURF. In *2011 International Conference on Computer Vision*. IEEE, Barcelona, Spain, 2564–2571. <https://doi.org/10.1109/ICCV.2011.6126544>
- [24] Amit Saxena, Mukesh Prasad, Akshansh Gupta, Neha Bharill, Om Prakash Patel, Aruna Tiwari, Meng Joo Er, Weiping Ding, and Chin-Teng Lin. 2017. A review of clustering techniques and developments. *Neurocomputing* 267 (Dec 2017), 664–681. <https://doi.org/10.1016/j.neucom.2017.06.053>
- [25] Christof Schöch. 2013. Big? Smart? Clean? Messy? Data in the Humanities. *Journal of Digital Humanities* 2, 3 (2013), 13 pages. <https://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/>
- [26] Kuo-Chung Tai. 1979. The Tree-to-Tree Correction Problem. *J. ACM* 26, 3 (Jul 1979), 422–433. <https://doi.org/10.1145/322139.322143>
- [27] Angela Weyer, Pilar Roig Picazo, Daniel Pop, JoAnn Cassar, Aysun Ozkose, Jean-Marc Vallet, and Ivan Srsa (Eds.). 2015. *EwaGlos - European Illustrated Glossary of Conservation Terms for Wall Paintings and Architectural Surfaces*. Petersberg, Germany. <http://openarchive.icomos.org/id/eprint/1706/>
- [28] Rui Xu and D. Wunsch. 2005. Survey of clustering algorithms. *IEEE Transactions on Neural Networks* 16, 3 (May 2005), 645–678. <https://doi.org/10.1109/TNN.2005.845141>