Automatic Structuring of Photographic Collections for Spatio-Temporal Monitoring of Restoration Sites: Problem Statement and Challenges

Laura Willot, D Vodislav, Livio De Luca, Valérie Gouet-Brunet

AUTOMATIC STRUCTURING OF PHOTOGRAPHIC COLLECTIONS FOR SPATIO-TEMPORAL MONITORING OF RESTORATION SITES: PROBLEM STATEMENT AND CHALLENGES

L. Willot [1,2,3], D. Vodislav [1], L. De Luca [2], V. Gouet-Brunet [3]
laura.willot@cyu.fr, dan.vodislav@cyu.fr, livio.deluca@map.cnrs.fr, valerie.gouet@ign.fr

[1] CY Cergy Paris Université, ENSEA, CNRS, ETIS, F-95000 Cergy, France
[2] MAP, UMR 3495, CNRS/MC, Marseille, France
[3] LASTIG, Univ. Gustave Eiffel, IGN-ENSG, 94160 Saint-Mande, France

Commission II

KEY WORDS: heritage science, computer vision, content-based retrieval, similarity metrics, multimodal data analytics, data semantics

ABSTRACT:

Over the last decade, a large number of digital documentation projects have demonstrated the potential of image-based modelling of heritage objects in the context of documentation, conservation, and restoration. The inclusion of these emerging methods in the daily monitoring of the activities of a heritage restoration site (context in which hundreds of photographs per day can be acquired by multiple actors, in accordance with several observation and analysis needs) raises new questions at the intersection of big data management, analysis, semantic enrichment, and more generally automatic structuring of this data. In this article we propose a data model developed around these questions and identify the main challenges to overcome the problem of structuring massive collections of photographs through a review of the available literature on similarity metrics used to organise the pictures based on their content or metadata. This work is realized in the context of the restoration site of the Notre-Dame de Paris cathedral that will be used as the main case study.

1. INTRODUCTION

Over the last decade, a large number of digital documentation projects have demonstrated the potential of image-based modelling (photo modelling, photogrammetry, ...) of heritage objects in the context of documentation, conservation, and restoration. The inclusion of these emerging methods in the daily monitoring of the activities of a heritage restoration site (context in which hundreds of photographs per day can be acquired by multiple actors, in accordance with several observation and analysis needs) raises new questions at the intersection of big data management, analysis, semantic enrichment, and more generally automatic structuring of this data.

Different methods are available to structure massive collections of photographs acquired in the aforementioned context. These methods often rely on different types of data used to enrich the photographs. Indeed, for the purpose of monitoring heritage sites, images are meaningful in more than one way: their visual content is as relevant as the spatial, temporal, and semantic dimensions expressed through the joint use of heterogeneous forms of metadata. Thus, the structuring of such large photographic collections implies the description, analysis, indexing, and classification of the images based on all diverse enriching data.

1.1 Case Study: Notre-Dame de Paris

The experiments that will be conducted during this project will take advantage of the massive collection of thousands of photographs taken before and during the restoration of the Notre-Dame de Paris cathedral.

Over several years, lots of pictures of the cathedral in various states have been taken by many different actors for diverse purposes. These photographs have been collected by the digital data working group (one of the nine working groups of the restoration project) and made available to all the actors of the scientific action for the restoration of the cathedral. The monitoring of the site can benefit from the study of these images as they carry significant information on the restoration process. For instance, after the fire, a lot of debris from the roof structure and vaults collapsed to the ground. To evaluate the damage and follow the clearance of spaces, several acquisitions have been carried out. In particular, several photographs of the pallets used to classify and collect the debris were taken, as shown on Figure 1. Knowing where and when these pictures have been taken, as well as their semantic content (the description of the objects collected) is an essential information to follow the life of the different debris from their discovery in the cathedral to their replacement or relocation. Therefore, these photographs are a direct and essential tool that will help the restoration process.

To provide the different actors with an access to the images, the digital data working group is setting up a complete digital platform for managing the entire data life cycle within a collaborative framework (https://www.notre-dame.science). This platform includes software bricks (or tools) for the human-driven description

(by metadata) and annotation of 2D images and related 3D models (such as a photogrammetric 3D reconstruction coming from the initial photographs, or any 3D model acquired by other means and directly linked to the images).

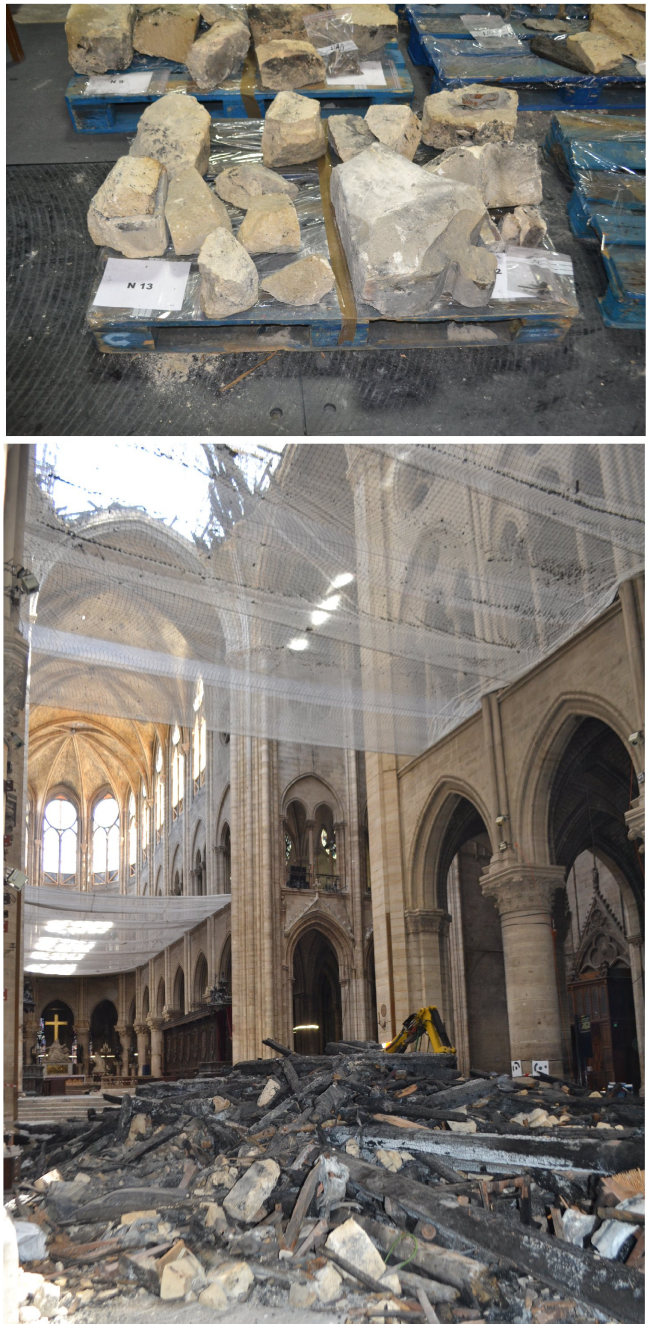

Figure 1: Examples of pictures taken in the context of a restoration site. **Top**: A pallet of debris (Claudine Loisel / LRMH) **Bottom**: The debris in the nave of the Notre-Dame cathedral (Aurélia Azéma / LRMH)
© Chantier Scientifique Notre-Dame de Paris / Ministère de la Culture / CNRS

Restoring a building, would it be as major as this cathedral or not, is a long-term process during which many new data will be acquired. The resulting collection of enriched photographs is a dynamic, ever-changing dataset: new pictures and new metadata are created every day.
However, enriching, documenting, and exploring this massive dataset of thousands of pictures are time-consuming tasks that are still mainly human-driven. An automatization of these tasks would nonetheless be possible through the structuring of the photographic collection. Thus, man-led enrichment would lead to automatic classification processes and ease the access to, and use of the information retained in the images and their metadata.

1.2 Aim and Structure

This paper aims to review the existing approaches that have been developed to solve the complex problem of structuring massive collections of photographs for the monitoring of restoration sites. First, a section on the data model will begin with an overview of the varied potential sources of data and data enrichment, available in a restoration project. This will be followed by a presentation of the data model itself in the current state of knowledge. Afterwards, we will develop on the different dimensions involved in heritage sites monitoring and how similarity measures are used to analyse and structure photographic collections. Then a discussion will be presented before a conclusion drawing on this review.

2. DATA MODEL

As explained above, for the structuring of massive collection of photographs, it is important to consider all heterogeneous data used to enrich the images themselves. The potential sources of data enrichment need to be identified in order to build the data model that will be the basis for future work.
In this section, we first describe the data available within a restoration project, exemplified by the work of the scientific action for the reconstruction of the Notre-Dame cathedral. Then, we present a primary version of the data model, built with regards to the data life cycle.

2.1 Data Life Cycle

Given the specific context of Cultural Heritage restoration sites monitoring, we define the data life cycle as the whole process the data is going through, from its acquisition to its use and re-use (e.g., through analysis, validation, or data enrichment). In practice, the ‘information object’ (the information carrier) is enhanced with layers of metadata associated through several methods, whether manual or automated (Gilliland, 2016). In the following paragraphs, we will look at the different sources of information rather than the knowledge acquired from them. This means that, for instance, even though temporality might have different finalities based on the needs, the focus will be put on the data itself (e.g., a time stamp in the image metadata) rather than what it means for a given actor.
For the structuring of photographic collections, the ‘information object’ to consider is the photograph. During their life cycle, many actors are involved to refine and develop on the knowledge

acquired beyond the observation of their plain visual aspects. For any restoration project, tools are developed to ease the access and storage of corpus of images : databases, thesauri and controlled vocabularies, 2D and 3D annotation tools, etc. Thus, the enrichment data can take various forms such as spatio-temporal or semantic metadata, 3D models, etc.

Spatio-temporal metadata: Spatio-temporal content is usually added in the metadata of the images at the very moment they are taken. Thus, localisation and time of capture are embedded in the image and can be directly accessed. However, it is also possible that the picture to analyse has been acquired through other means (e.g., digitisation of a physical photograph) and spatio-temporal metadata might not be available. Thus, it is important to acknowledge the potential incompleteness of metadata, whether spatio-temporal or not.

Semantic metadata: Semantic metadata usually incorporate annotations based on controlled vocabularies, which provide "standardized words and phrases used to refer to ideas, physical characteristics, people, places, events, subject matter, and many other concepts" (Harpring, 2010). Semantic tags can be added by different actors whose roles, functions or areas of expertise will impact their choice of word and result in a multi-disciplinary description of the image. This multi-layered vision has also an important role in the structuring of the collection of photographs as any classification of the photographs is influenced by the professional who is looking for specific information within the collection.

3D content: Modern data enrichment and structuring can also be done through the association of photographs and 3D content such as 3D models of the scene, enriched through various methods. It can take the form of point clouds which result from photogrammetric algorithms or 3D scanning (LiDAR), or of diverse 3D models of monuments or territories such as those provided by mapping agencies. Generally, they are highly structured and geolocalized, and then can serve as a common ground for structuring the photographic collections (Blettery et al, 2021). They can be associated, or enriched later on, with reality-based 3D annotations, a hybrid type of annotation used to add information on both the 2D and the 3D content (Manuel et al., 2016). These 2D/3D annotations are another important source of semantic knowledge and could be interrogated to get more information on the pictures, thanks to the continuous projective relationship established between the images and the 3D model enabling the propagation of the information from 2D to 3D medium and vice versa. Other techniques for point clouds exploration can include machine learning based ones (Poux et al., 2016).

Visual content: Beside the adjunction of metadata to the images, one of the main sources of information remains the image itself. The visual content of the photograph is of utmost importance as the primary carrier of information. As a matter of fact, the production or rather acquisition of a photograph is primarily done to save a spatio-temporal unit of visual information.

The automatic analysis of the visual content of an image is a problem which has been widely studied since several decades, with many varied applications such as pattern recognition, classification, image indexing and retrieval, photogrammetry, computer vision and robotics. They share the common objective of extracting from the image content an information that will allow its semantic interpretation, its comparison with other contents (e.g. for 3D reconstruction), its indexing in a collection or interlinking with other contents.

2.2 Data Model

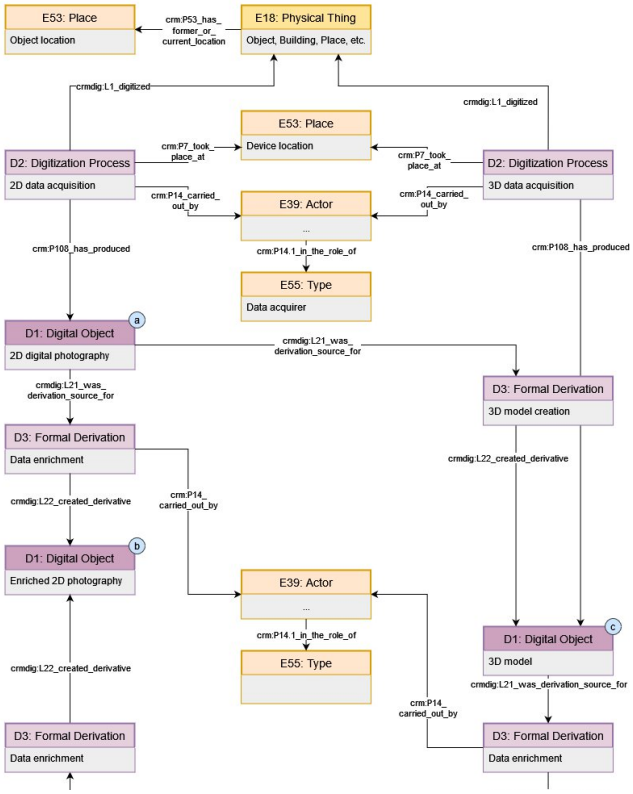

Figure 2: An initial version of the Data Model representing the various links between the digital objects considered, whether 2- or 3-dimensional. CIDOC-CRM classes are in yellow and CRMdig ones in purple. Letters **a** to **d** refer to the images provided as examples in Figure 3

Our data model is based on the CIDOC Conceptual Reference Model (CRM), a "formal ontology intended to facilitate the integration, mediation and interchange of heterogeneous cultural heritage information and similar information from other domains" (Bekiari et al., 2021). It has been developed over two decades by the International Committee for Documentation (CIDOC) of the International Council of Museums (ICOM). It became an ISO standard in December 2006 and is in its latest version since April

2021 (Bekiari et al., 2021). Its main focus was to allow exchange and sharing of information within a harmonised framework.

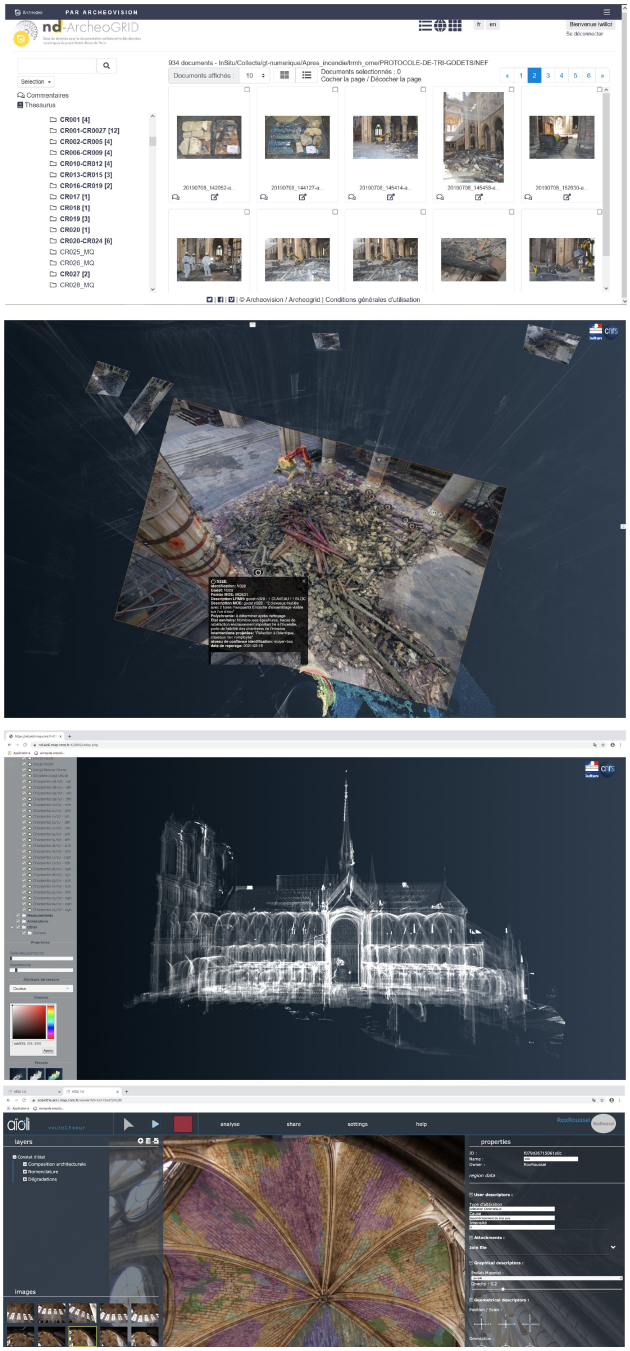

Figure 3: Illustrations of the different Digital Objects presented in the Data Model. From top to bottom: **a.** Image gallery of the photographic collection (Sarah Tournon / Archeovision) **b.** Projection of spatially oriented and annotated picture onto the point cloud of the Notre-Dame cathedral (Violette Abergel, Livio De Luca / MAP / Vassar College / SRA-DRAC / AGP / LRMH) **c.** 3D point cloud of the cathedral (Violette Abergel / MAP / Vassar College) **d.** 2D/3D annotations of the alterations found on the vaults of the choir (Roxane Roussel / MAP)
© Chantier Scientifique Notre-Dame de Paris / Ministère de la Culture / CNRS

The first versions of the CIDOC-CRM were built on an Entities-Relationships model until the CIDOC decided to adopt an object-oriented approach (Oldman, 2014). This resulted in the CIDOC-Conceptual Reference Model, easier to extend for specific use-cases. Subsequently, several extensions have been proposed – and adopted – to allow Cultural Heritage professionals to define more specific entities and properties, in order to be more accurate for the particular cases considered. Among the existing extensions of the CIDOC-CRM, we can cite for instance: CRMcr for the conservation and restoration domain (Bannour et al., 2018), CRMgeo dedicated to spatio-temporal contextualisation of the data (Migliorini, 2018; Nys et al., 2018), and CRMdig for digital elements (Doerr et al., 2016). Each of these extensions expresses different needs that occurred in different situations. They can then be used to build conceptual models that reflect the specific requirements of any given situation in which the CIDOC-CRM was not precise enough.

As a major tool increasingly used by Cultural Heritage professionals and in order to follow a shared understanding of information, we decided to use the generic CIDOC-CRM and its digital extension, CRMdig, to build our data model. These two sets of classes contained the most relevant entities to represent the data identified in the previous subsection. The resulting data model we propose is represented in Figure 2. The CRMdig classes are coloured in purple and the CIDOC-CRM ones in yellow.

Based on the case study, we identified various sources of information, both textual and visual, that were associated with the images throughout their life cycle. Each of these information is the result of an activity performed by a given actor on a specific (digital or physical) object. The main distinction we made between the digital objects derive from their dimensionality. On the left side are represented the acquisition and enrichment processes of 2D content while 3D content production and annotation is on the right side. Both content are related to the same physical object. The letters a to d refer to the illustrations in Figure 3 used to exemplify the different Digital Objects of the Data Model.

3. PHOTOGRAPHIC COLLECTIONS STRUCTURING

Structuring a massive collection of photographs is a problem that finds its roots in the difficult question of organising this data by similarities. Indeed, specific characteristics related to the images content, or to their metadata can be identified and used to describe and compare pictures, which opens the way to other more advanced analyses such as interpretation, indexing and reconstruction. The resulting links between the elements compared together can then be represented within a network of relations: each image is represented by a graph node, whereas the edges depict the existence of a similarity between nodes/images.

But, as exposed in the previous section, structuring a collection of photographs for monitoring purposes implies various sources of

information used to enrich the visual content of the pictures. Similarity thus becomes an equivocal term depending on a given criterion used compare images along a specific dimension. The subsequent connections between the images would then be represented in a multi-layered network of relations, allowing multi-criteria search and exploration.

The spatio-temporal monitoring of Cultural Heritage restoration sites relies on three main dimensions: time, space, and semantics. Indeed, when restoring, spaces can quickly evolve. Taken at different stages, the pictures can then serve as testimonies of how the site looked like at a given moment, or where elements were located for a specific period of time and what they represent.

In this section, we will review the literature dealing with similarity metrics used to measure a 'distance' between images, along the three aforementioned dimensions. Given the heterogeneity of the information brought during the data enrichment process, diverse techniques and methods will be considered, whether the data to evaluate is of a textual or visual nature.

3.1. Time

Time is one of the fundamental dimensions necessary to the monitoring of restoration sites. Knowing when objects have been moved or how a place has evolved during the process are ones of the many time-related tasks performed by the diverse actors implied in such a project.

The temporal dimension is twofold: on one side, one photograph represents a short unit of time, the precise moment it was taken, and on the other side, the joint study of several photographs allow a visualisation with a larger time span.

Thus, temporal similarities between photographs can be studied through both the metadata (e.g., a timestamp indicating the date and time of the image capture) and the visual content of the pictures. Depending on the targeted objective, one ambition will be to develop tools that are sensitive to this criterion of time, or on the contrary, invariant to it (e.g., connecting photographs of objects taken at different time periods), which will have a deep impact on the nature of the description techniques based on visual appearance analysis.

3.2. Space

Often related to temporal analyses, space is another important aspect for the monitoring of a restoration site. Similarly to the temporal dimension, two aspects of space are equally meaningful: on one hand, it is interesting to know where the cameras were located when used to take the pictures, and on the other hand, the visual content of the images depicts items that are potentially moveable and which moves need to be registered (see Figure 1 for an example of moveable elements).

Among topical applications of automatic analysis of an image visual content, we can mention the problem of visual-based localization whose objective is to geolocalize the content (or the camera), which is usually processed in an unsupervised way with content-based image retrieval tools (Pion et al, 2020) (Blettery et al, 2021). Other techniques include spatial reasoning based on computer vision and description logic (Hudelot et al., 2014).

Another significant spatial information source comes from the strong connection between 2D and 3D content we highlighted in our data model. Spatial information can be transferred from one medium to the other or presented simultaneously in immersive hybrid viewers or web platforms (Stefani et al., 2013; Paiz-Reyes, 2019). Figure 4 illustrates a projection of a spatially oriented picture onto a point cloud of the Notre-Dame cathedral and visualised in a 3D viewer. This type of projection within a 3D environment gives a better understanding of the spatialization of the pictures.

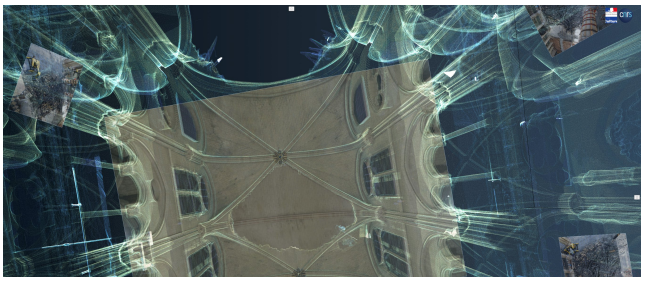

Figure 4: Projection of a spatially oriented photograph onto the point cloud of the Notre-Dame de Paris cathedral displayed within the 3D viewer developed by the MAP laboratory as part of the scientific action for the reconstruction of the Notre-Dame de Paris cathedral. (Violette Abergel, Livio De Luca)

Spatial proximity between images is a more complex problem than the mere comparison of devices' localizations and both the camera and depicted objects' localizations need to be acknowledged.

3.3. Semantic aspect

As explained before, semantic metadata usually take the form of terms chosen among a list defining a controlled vocabulary, an "organized arrangement of words and phrases used to index content and/or to retrieve content through browsing or searching" (Harpring, 2010). An example of hierarchical controlled vocabulary is given in Figure 5 with a section of the tree structured thesaurus used for the Notre-Dame restoration project; each of these terms might be associated with any image of the dataset. This semantic information can be used for exploration and retrieval purposes, but here we focus on measuring the similarity between images based on the proximity of the terms used to describe the pictures to compare. One way to build such semantic metrics is to measure this proximity between words based on shared information content (taxonomic links expressed through structured hierarchical graphs) and probability estimates (Resnik, 1999).

However, the semantic description of a picture can include several such term-based annotations, but also numerical properties. This multi-property semantic characterization may take the form of RDF semantic graphs, where each object (picture) is a graph node

and its properties are oriented graph edges starting from that node. Measuring the semantic similarity between pictures in this context comes down to measure the similarity between the corresponding nodes and their neighbourhood in the semantic graph. This kind of graph similarity metrics is used e.g. for graph querying (De Virgilio et al., 2013), where one measures the similarity between a small query graph and parts of the a larger RDF graph (representing the dataset on which the query is executed) by looking for the portions of the larger graph most similar to the query. Similar metrics are used in the field of ontology matching (Shvaiko and Euzenat, 2013; Liu et al., 2021), where the aim is to handle the problem of semantic heterogeneity, by finding relations between entities from distinct ontologies that share a similar description. Another possibility to assess the similarity of semantic graph nodes is to use graph embeddings (Trisedya et al., 2019) that map graph nodes into points in a high dimensional metric space, and measure node similarity through the distance between the corresponding points.

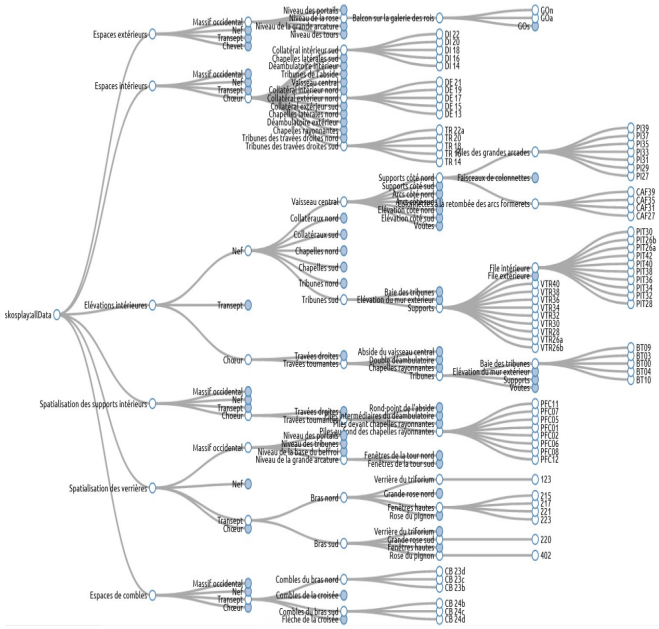

Figure 5: A fragment of the Notre-Dame thesaurus tree structure, built by the digital data working group, encoded in SKOS via OpenTheso, a thesaurus manager.
© Miled Rousset / MOM / Isabelle Cao / MAP / Denis Hayot / PLEMO 3D / SKOS Play / SPARNA / Chantier Scientifique Notre-Dame de Paris / Ministère de la Culture / CNRS – 2021

Although presented as separate "entities", time, space, and semantics can be studied separately or jointly, depending on the needs. There also exist common lower level processing bricks allowing to extract information along these axes. For instance, visual content analysis can be considered as a low-level processing thanks to which new temporal, spatial or semantic knowledge is created, in particular when metadata are lacking on a particular aspect. Literature on the automatic description of visual contents and their comparison (visual similarity) is very large, with several surveys focusing either on recent powerful approaches based on deep learning (Ma et al., 2021) or on the exploitation of other available modalities to improve description and matching (Piasco et al., 2018). Among the remaining challenges in this area, specialized or heritage iconography datasets are obviously under-represented in the multiple datasets exploited with deep-based descriptions, conducing to the development of the few-shot learning branch, which focuses on the learning of descriptions using a limited number of annotated data (Wang et al., 2020). Figure 6 illustrates the problem of the visual description, matching and indexing of heritage iconography of landscapes with application to their interlinking and geolocalization. Here the description learnt allows to focus on the spatial axis by retrieving similar locations, while omitting the temporal axis by retrieving images at different time periods.

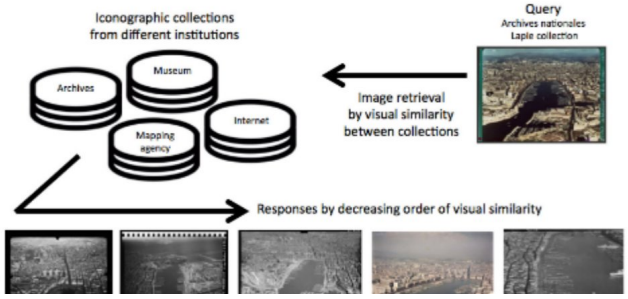

Figure 6: Example of content-based image interlinking across collections exploiting visual similarity, dedicated to the retrieval of location instances at different time periods, here Marseille's port (France) (Gominski et al, 2021)
Image by courtesy of Alegoria project: https://www.alegoria-project.fr

4. DISCUSSION

The project we present here intends to address the complex issue of automatic structuring of photographic collections, with regards to the temporal, spatial and semantic dimensions of such pictures. As a central element to perform indexing and classification tasks, the structuring of photographic collections will allow a better handling and valorization of these collections for the many actors who use these pictures as an extensive source of knowledge.
With a direct application to the scientific action for the restoration of the Notre-Dame de Paris cathedral, this project will deal with the complex and constantly evolving database of several thousands of pictures. Thus, the data model we presented here has been based on the data and tools available for this precise restoration project. Nonetheless, the methodological approach that we will develop is expected to be reproducible and general enough to be of use for other heritage restoration projects.
To perform corpus exploration, we aim to build a multi-layered network representing the links between the images, based on their similarities in terms of space, time and semantics. We have defined these three aspects as the main characteristics used for restoration site monitoring ; they can be expressed in several criteria

representing in a multidimensional way the similarities between the photographs. To build this graph, we aim to define multi-criteria similarity metrics.
Perspectives of our work include higher-level tasks performed on the graph, such as multi-criteria search, content retrieval, or propagation of data annotations from picture to picture.

5. CONCLUSION AND FUTURE WORK

Following the advent of new digital techniques such as AI-based image structuring , image-based modelling, and other computer vision-related techniques, cultural heritage projects have been producing an increasing number of pictures taken daily that can be used to follow-up the monitoring of restoration processes.
This paper presents the first results and the work that will be conducted towards an automatic structuring of those collections of photographic images, with a direct application to the restoration of the Notre-Dame de Paris cathedral. These results include an initial inventory of the main types of image enriching content. This inventory has been based on the tools available for the Notre-Dame project with regards to the data life cycle that each image is going through, from its acquisition to its use and re-use by the different actors of the project. The current literature on similarity measure techniques, in accordance with temporal, spatial and semantic dimensions, has also been reviewed.
Only preliminary work has been presented here and will be expanded in a near future. Future work include a better definition of the data model which will become the source material for the development of a methodological approach to content-based image comparison. This methodology will consider the spatial, temporal, and semantic dimensions involved in the life cycle of these photographs, which have been defined as the main aspects for monitoring the restoration process. This should be translated by the creation of similarity metrics used to build a multi-layered network of relations representing the connections between the images. Thereupon, this network might be used to perform automatic high-level tasks, useful to the many actors involved with the photographic images acquired during the restoration process.

ACKNOWLEDGEMENTS

This work is supported by the CYU Graduate School Humanities, Creation, Heritage, Investissement d'Avenir ANR-17-EURE-0021 – Foundation for Cultural Heritage Science

REFERENCES

Bannour, I., Marinica, C., Bouiller, L., Pillay, R., Darrieumerlou, C., Malavergne, O., Kotzinos, D., Niang, C., 2018. CRMcr - a CIDOC-CRM extension for supporting semantic interoperability in the conservation and restoration domain, in: Digital Heritage 2018. San Francisco, United States.

Bekiari, C., Bruseker, G., Doerr, M., Ore, C.-E., Stead, S., Velios, A. (Eds.), 2021. Definition of the CIDOC Conceptual Reference Model (Version 7.1.1 Produced by the CIDOC CRM Special Interest Group).

Blettery, E., Fernandes, N., Gouet-Brunet, V., 2021. How to Spatialize Geographical Iconographic Heritage, in: Proceedings of the 3rd Workshop on Structuring and Understanding of Multimedia heritAge Contents (SUMAC 2021 @ ACM Multimedia 2021), Oct 2021, Chengdu, China.

Chantier Scientifique Notre-dame de Paris, Ministère de la Culture / CNRS, 2020. URL https://www.notre-dame.science/ (accessed 1.20.22).

De Virgilio, R., Maccioni, A., Torlone, R., 2013. A similarity measure for approximate querying over RDF data, in: Proceedings of the Joint EDBT/ICDT 2013 Workshops on - EDBT '13. Presented at the the Joint EDBT/ICDT 2013 Workshops, ACM Press, Genoa, Italy, p. 205. https://doi.org/10.1145/2457317.2457352

Doerr, M., Stead, S., Theodoridou, M., 2016. Definition of the CRMdig An Extension of CIDOC-CRM to support provenance metadata 18.

Gilliland, A. J., 2016. Setting the Stage, in: Baca, M. (Ed.), Introduction to Metadata, Introduction To. Getty Publications.

Gominski, D., Gouet-Brunet, V., and Chen, L., 2021. Connecting Images through Sources: Exploring Low-data, Heterogeneous Instance Retrieval, Remote Sensing Journal (MDPI), Special Issue "Digitization and Visualization in Cultural Heritage", Vol 13, no. 16: 3080, https://doi.org/10.3390/rs13163080

Harpring, P., 2010. Introduction to Controlled Vocabularies: Terminology for Art, Architecture, and Other Cultural Works, Online Edition. ed, Introduction To. Getty Research Institute, Los Angeles, CA.

Hudelot, C., Atif, J., Bloch, I., 2014. ALC(F): A new Description Logics for Spatial Reasoning in Images, in: 1st International Workshop on Computer Vision + ONTology Applied Cross-Disciplinary Technologies (CONTACT 2014). Zurich, Switzerland, pp. 370–384.

Liu, Xiulei, Tong, Q., Liu, Xuhong, Qin, Z., 2021. Ontology Matching: State of the Art, Future Challenges, and Thinking Based on Utilized Information. IEEE Access 9, 91235–91243. https://doi.org/10.1109/ACCESS.2021.3057081

Ma, J., Jiang, X., Fan, A., Jiang, J., Yan, J., 2021. Image Matching from Handcrafted to Deep Features: A Survey. Int J Comput Vis 129, 23–79. https://doi.org/10.1007/s11263-020-01359-2

Migliorini, S., 2018. Enhancing CIDOC-CRM Models for GeoSPARQL Processing with MapReduce, in: Belussi, A., Billen, R., Hallot, P., Migliorini, S. (Eds.), Proceedings of the 2nd Workshop On Computing Techniques For Spatio-Temporal Data in Archaeology And Cultural Heritage, CEUR Workshop Proceedings. Presented at the Computing Techniques For Spatio-

Temporal Data in Archaeology And Cultural Heritage, CEUR, Melbourne, Australia, pp. 51–65.

Nys, G.-A., Ruymbeke, M.V., Billen, R., 2018. Spatio-Temporal Reasoning in CIDOC CRM: An Hybrid Ontology with GeoSPARQL and OWL-Time, in: Belussi, A., Billen, R., Hallot, P., Migliorini, S. (Eds.), Proceedings of the 2nd Workshop On Computing Techniques For Spatio-Temporal Data in Archaeology And Cultural Heritage, CEUR Workshop Proceedings. Presented at the Computing Techniques For Spatio-Temporal Data in Archaeology And Cultural Heritage, CEUR, Melbourne, Australia, pp. 37–50.

Oldman, D., 2014. The CIDOC Conceptual Reference Model (CIDOC-CRM): PRIMER 22.

Paiz-Reyes, E., 2019. Image based rendering of large historical image collections, in: Computer Graphics Forum. Genova, Italy.

Piasco, N., Sidibé, D., Demonceaux, C., Gouet-Brunet, V., 2018. A survey on Visual-Based Localization: On the benefit of heterogeneous data. Pattern Recognition 74, 90–109. https://doi.org/10.1016/j.patcog.2017.09.013

Pion, N., Humenberger, M., Csurka, G., Cabon, Y., Sattler, T., 2020. Benchmarking Image Retrieval for Visual Localization, In International Conference on 3D Vision, 2020.

Poux, F., Hallot, P., Neuville, R., Billen, R., 2016. Smart Point Clouds: Definition and Remaining Challenges, in: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Presented at the TC IV 11th 3D Geoinfo Conference, Copernicus GmbH, Athens, Greece, pp. 119–127. https://doi.org/10.5194/isprs-annals-IV-2-W1-119-2016

Resnik, P., 1999. Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language. jair 11, 95–130. https://doi.org/10.1613/jair.514

Shvaiko, P., Euzenat, J., 2013. Ontology Matching: State of the Art and Future Challenges. IEEE Transactions on Knowledge and Data Engineering 25, 158–176. https://doi.org/10.1109/TKDE.2011.253

Stefani, C., Busayarat, C., Lombardo, J., Luca, L.D., Véron, P., 2013. A web platform for the consultation of spatialized and semantically enriched iconographic sources on cultural heritage buildings. J. Comput. Cult. Herit. 6, 1–17. https://doi.org/10.1145/2499931.2499934

Trisedya, B. D., Qi, J., Zhang, R., 2019. Entity alignment between knowledge graphs using attribute embeddings. AAAI 2019, Article 37, 297–304. https://doi.org/10.1609/aaai.v33i01.3301297

Wang, Y., Yao, Q., Kwok, J.T., Ni, L.M., 2020. Generalizing from a Few Examples: A Survey on Few-shot Learning. ACM Comput. Surv. 53, 63:1-63:34. https://doi.org/10.1145/3386252