Science Arts & Métiers (SAM)
is an open access repository that collects the work of Arts et Métiers Institute of Technology researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: https://sam.ensam.eu
Handle ID: .http://hdl.handle.net/10985/17575

To cite this version :

Violette ABERGEL, Kévin JACQUOT, Livio DE LUCA, Philippe VERON - Towards a SLAM-based augmented reality application for the 3D annotation of rock art - International journal of virtual reality - Vol. 19, n°2, p.nc - 2019

Any correspondence concerning this service should be sent to the repository Administrator : scienceouverte@ensam.eu

Towards a SLAM-based augmented reality application for the 3D annotation of rock art

Violette Abergel [1], Kévin Jacquot [1], Livio De Luca [2], Philippe Veron [3]

[1] MAP-Aria, UMR 3495 CNRS/MC, Lyon, France
[2] MAP-Gamsau, UMR 3495 CNRS/MC, Marseille, France
[3] LISPEN, EA 7515 ENSAM, Aix-en-Provence, France

Abstract - **The digital technologies developed in recent decades have considerably enriched the survey and documentation practices in the field of cultural heritage. They now raise new issues and challenges, particularly in the management of multidimensional datasets, which require the development of new methods for the analysis, interpretation and sharing of heterogeneous data. In the case of rock art sites, additional challenges are added to this context, due to their nature and fragility. In many cases, digital data alone is not sufficient to meet contextualization, analysis or traceability needs.**
In this context, we propose to develop an application dedicated to rock art survey, allowing 3D annotation in augmented reality. This work is a part of an ongoing project about an information system dedicated to cultural heritage documentation. For this purpose, we propose a registration method based on a spatial resection calculation. We will also raise the perspectives that this opens up for heritage survey and documentation, in particular in terms of visualization enhancement.

Index Terms - **Augmented reality; Cultural heritage survey; Rock art; Semantic annotation; WebXR.**

I. Introduction

The work described in this paper focuses on the field of cultural heritage survey, and more specifically applied to the corpus of French rock art sites. Recent decades have been characterized by the development of digital technologies, which have produced spectacular advances in the gathering, viewing and indexing of digital resources. While the new tools introduced through these advances have enabled a significant change in documentation practices within the cultural heritage community, they now raise new issues and challenges. In particular, multidimensional and multiformat data management involves issues related to the development of new methods of analysis and interpretation, the sharing and correlation of heterogeneous data across multiple actors and contexts, and the centralized archiving of documentation results for long-term preservation purposes.

E-mail: violette.abergel@lyon.archi.fr; kevin.jacqot@lyon.archi.fr; livio.deluca@map.cnrs.fr; Philippe.Veron@ensam.eu

Moreover, the democratization of these digital tools led to an important paradigm shift in the field of heritage survey, usually considered as a process consisting of three distinct phases: “description”, “figuration” (interpretation), and “representation” (Saint Aubin 1992). The interpretation step is more and more frequently performed with a digital clone, and is thus being remotely handled behind a computer rather than in direct contact with the real object. In the case of rock art, it raises various concerns and challenges, because the extreme fragility of the sites, their characteristics, and the multidisciplinary nature of the teams mobilized for their study and monitoring impose a number of precautions to:

– minimize the data scattering,

– facilitate cross-analysis of observations,

– ensure the scientificity/reproducibility of the data.

Thus, according to the CNP (the French National Center of Prehistory), it becomes even more imperative to ensure the contextualization, temporalization and traceability of data relating to each object of study, but also to propose new analytical tools allowing the visual accentuation of painted or engraved elements, which are increasingly difficult to perceive due to the natural degradation of the caves. The key therefore seems to be the reunification of the three survey phases into a single moment in direct contact with the real object. For this reason, augmented reality presents a strong potential in several respects: on the one hand to become an interface between the essential in-situ presence and the digital instance - allowing work to be done on the virtual copy when the context permits it - and on the other hand to set up tools for visual enhancement of perceptual attributes, facilitating experts’ observation work.

In this context, our approach consists in an augmented reality web application dedicated to rock art survey, allowing the visualization and saving of any observation, measurement, analysis or data in its real context. More specifically, the aim is to extend the functionalities of a pre-existing heritage information system Aïoli (De Luca et al. 2018) by adapting it to the constraints related to cave studies and by proposing a semi-automatic method of real/virtual alignment based on photogrammetry, which we will describe in greater detail in this paper.

II. Context: the rock art survey

For this project, we focus on the corpus of rock art sites on French territory. This covers about 180 sites (caves or shelters either decorated, engraved or sculpted), the majority of which are classified as Historical Monuments. These sites are extremely fragile and do not support any form of restoration. As a result, their access is regulated and their study by scientific teams increasingly requires their digitization to deport part of the work when possible. However, round trips to the real object or to the facsimile if it exists remain essential.

The archaeological survey applied to rock art includes two main methods which differ in the weighting given to the three phases of the survey. The plastic analytical survey, which responds to two concerns to analyze and make comprehensible includes a large part of "figuration". The technical analytical survey, recommended by the CNP, which corresponds to an exhaustive mapping of the elements of the walls of the cave studied, includes a dominant "description". In both cases, the link to the real object remains essential: although some graphical parts of the survey are carried out offsite through the data collected, both methods involve many round trips to the real object to compare the survey with what can be observed or perceived in situ (Fuentes 2017).

The rock art study involves a wide variety of actors and disciplines, including archaeology, anthropology, palaeogenetics, geology, climatology, microbiology, etc. These actors each mobilize various methods and tools, which naturally leads to a dispersal effect of data and makes it difficult to compare the observations and analyses carried out within these multidisciplinary teams. The identification of a stable common denominator therefore appears to be an essential prerequisite for collecting, organizing, safeguarding and comparing the various contributions. Augmented reality appears in a second stage as a serious avenue to allow the capture, contextualized visualization and traceability of all content.

III. Related works

To address these challenges, current research is focusing partly on the implementation of information systems dedicated to cultural heritage, and secondly on the creation of augmented reality content, mainly for dissemination purposes.

1. Information systems and cultural heritage

Knowledge-centered approaches aim to improve content data management by defining formal structures that describe the implicit and explicit relationships of concepts linking elements together (Doerr 2009), combined with tools for enriching documents by linking them to semantic tags (Havemann et al. 2009; Manferdini et al. 2008; Myers et al. 2012; Wefers et al. 2016). But reality-based 3D reconstruction and 3D information systems are generally considered as distinct topics and few works consider them as complementary parts of a single framework, even if they use very different approaches. On the one hand, knowledge-centered information systems are being used for the management of documentation activities (Blaško et al. 2012; Boochs et al. 2014), but without explicit or direct exploitation of the geometric and visual properties of 3D data. On the other hand, some 3D-centered approaches allows the annotation of reality-based models (Apollonio et al. 2018; Shi et al. 2016; Soler, Melero, and Luzón 2017) but misses any reference to the complex relationships inherent to the documentation and interpretation process.

Finally, (Manuel et al. 2018) focuses on merging geometric, visual and semantic aspects into a single integrated documentary approach, with the objective of offering through a cloud platform an information continuum that merges the acquisition, analysis, and interpretation phases in a collaborative framework. This approach relies on an indexing phase following the photogrammetric process, which establishes a 2D/3D projective relationship by keeping in memory the link between the pixels of the different photographs in a dataset and the points of the resulting point cloud. This process is then used to allow the propagation of annotations on all 2D or 3D resources belonging to a project (see section III.1).

Nevertheless, this approach does not currently allow neither a real-time management, which is however required for its use in augmented reality, nor the contextualization of data on reality. The link between 2D and 3D is established, the 3D model is well reality-based, but the 3D/reality link remains absent. This is a major gap that affects the traceability of user actions. Our study is an extension of this work, with the objective of extending the above-mentioned projective relationship to the real object, in order to allow the development of tools dedicated to augmented reality surveying. First of all, this implies resolving the problem of the coherent registration of Aïoli project data on reality. The following sections present a method for this point and discuss its results.

2. Augmented reality and cultural heritage

In the field of cultural heritage, augmented reality applications mainly concern mediation and dissemination. These applications, mainly for tourism or virtual museums (Bekele et al. 2018), require prior preparation of the 3D model, often in known or controlled environments (Clini et al. 2014; Damala et al. 2016; Madsen and Madsen 2015; Ridel et al. 2014). Such applications rarely use semantics, and in these rare cases, 3D data are processed and segmented upstream in order to define each interaction modality and scenarize the contents. These strategies cannot be suitable for us, since the goal of extending the acquisition - interpretation - representation continuum to the real object requires a certain "transparency" or genericity in the preparation of the data

If all the data of a project can be represented within a single space (see section III.3), then the main obstacle is AR registration. Existing work in this area includes different approaches. Electromagnetic or ultrasonic localization systems are not suitable in our case, since they require the environment to be adapted, which is precluded in the case of decorated caves due to the nature of the sites and their fragility. In the same way, the use of markers to improve the accuracy

of a vision-based localization is excluded. Moreover, localizations methods using exclusively an inertial sensor are not sufficient because they drift over time (Rolland, Baillot, and Goon 2001; Marchand, Uchiyama, and Spindler 2016), and location-based approaches using GNSS data would require ensuring the systematic geolocation of projects, which is not the case and does not seem appropriate because of the multiplicity of scales processed. Finally, markerless localization methods based on monocular SLAM (Taketomi, Uchiyama, and Ikeda 2017), sufficiently precise and generic but also inexpensive, could be adapted to our objectives. However, they have several drawbacks: on the one hand, a measurement drift observed over time, linked to error accumulation within the incremental process (Arnaldi, Guitton, and Moreau 2018), which can affect the accuracy of alignment, and on the other hand, in the lack of geolocation of images, the resulting poses are expressed only in a local reference system usually arbitrarily chosen from the initial camera reference system which obviously differs from the project one.

In addition, these methods, which usually require the use of a compiled and autonomous application, are gradually becoming exploitable on the web thanks to the emergence of new APIs, in particular WebVR (W3C 2017), which is used as a basis for the first experiments in producing augmented reality content in a browser (Bergstrom 2018; Medley 2018), and more recently the WebXR device API (W3C 2018). These drafts are combined with experimental browsers built on top of the Google© or Apple© AR SDK (Gosalia 2018; Buerli and Misslinger 2017). As a matter of fact, despite their experimental aspect, we are focusing our work on these environments in order to offer a cloud platform dedicated to augmented reality surveys. We assume that the implementation of SDKs dedicated to the extraction of camera poses and feature points is likely to improve over time as web standards evolve, and that the underlying algorithms used for SLAM could always be improved later and independently if our proof of concept were to be conclusive.

III. Main approach

In the following sections, we first propose a method allowing the unification of SLAM localization methods with Aïoli project data, in order to align 2D and 3D data in to their real context in a coherent way, without any distinction of content. Then, we focus on the implementation of dedicated survey tools to facilitate analysis and the prospects this opens up in terms of traceability and practices.

1. Aïoli

Aïoli (De Luca et al. 2018) is a cloud platform dedicated to 2D and 3D semantic annotation for collaborative documentation of heritage objects. Accessible from a browser on a PC, tablet or smartphone, it allows users to make spatialized 3D annotations from simple photographs around a study object (whether a building, a sculpture, a painting, a piece of art or archaeological fragments), and to make them available to their peers. It includes:

- An image-based incremental 3D spatialization process to manage the geometric fusion of multiple images from different users at different times;
- A 2D/3D annotation framework allowing users to draw, visualize and save relevant regions by manipulating simple spatially oriented 2D images around a dynamic 3D representation;
- A multi-layer morphological data structuring model to accurately describe real objects in all their geometric complexity and according to multidisciplinary observations.

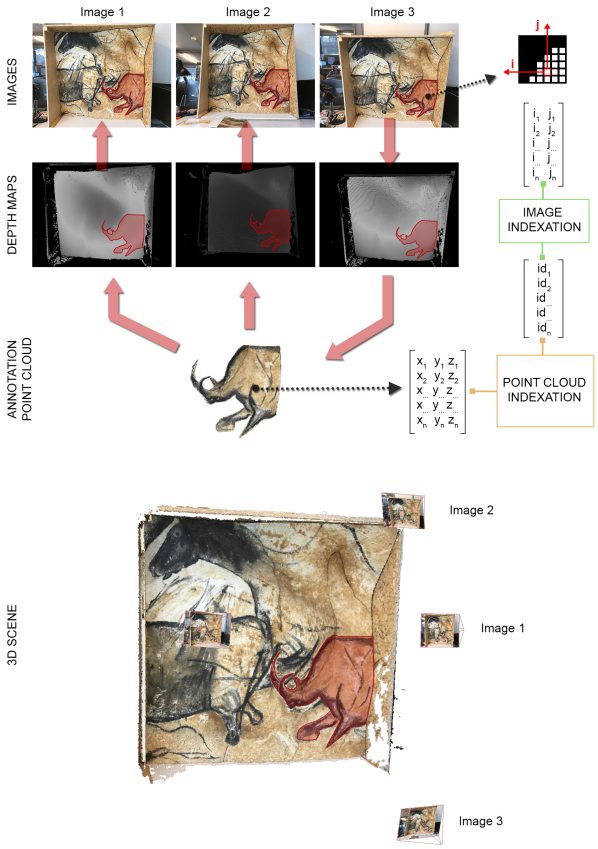

Figure 1: The indexing process, establishing a projective 2D/3D relationship

Based on the automation of image spatialization processes by photogrammetry and the ability to collect, process and distribute large amounts of data via cloud computing, the core of the platform is a multidimensional correlation engine. The photographs sent by each user during the creation of a project are used to generate a dense 3D point cloud of the object studied. By completing this process with an indexing phase, we establish a correspondence between each point of the point cloud and each pixel of the different images. Thus, this bijective relationship makes it possible to get, for each 3D point, the list of images on which the point appears. Conversely, starting from one or more points on an image, we can obtain the list of the 3D points concerned.

Thanks to this process, annotations made on any image of the object are automatically projected on all the other images (past, present and future) (Figure 1), but also continuously correlated geometrically, visually and semantically, with other annotations having a close spatial location. The annotations are structured as regions belonging to layers, associated with geometric (automatically calculated) and semantic (defined by the various experts) descriptors. In addition to the description fields, users can associate additional multimedia resources to these, which are then located in 3D on the barycenter of the annotation concerned.

The MicMac framework (Rupnik, Daakir, and Pierrot Deseilligny 2017) used for photogrammetry, allows to estimate the pose of an additional image from a set of already oriented images. In Aïoli, this feature allows to add one or more images to an existing project while retrieving annotations already made. For this purpose, two methods (Bundle Adjustment or Spatial Resection) have been integrated and can be used depending on the presence or absence of useful metadata (e.g. focal length) (Pamart, Morlet, and de Luca 2019).

2. Overall approach for AR registration

Assuming that SLAM allows us to estimate the camera pose in real time in a non-georeferenced arbitrary reference system, and that on the other hand, all the 2D and 3D geometric and semantic data of an Aïoli project can be represented in a single coherent scene that is also non-georeferenced, we suppose that the spatialization of one frame of the video stream of the device would allow us to determine its pose in the two spatial references, and thus to align the project data in a coherent way in relation to the camera stream (Figure 2).

A T_0, i.e., when creating a project, the user uploads his photographs, which are processed on a server by a spatial referencing process based on photogrammetry and the indexation of the resulting data. At the end of this process, the project includes a point cloud and oriented indexed images, all represented in an arbitrary reference system. The user must then manually set the scale using a raycast tool to pick two points from the point cloud and indicate the distance between them. At the end of this step, the project's 3D database is ready for annotations but also for augmented reality.

At each T_n, i.e. at each use session of a project, the augmented reality visualization mode can be activated from the

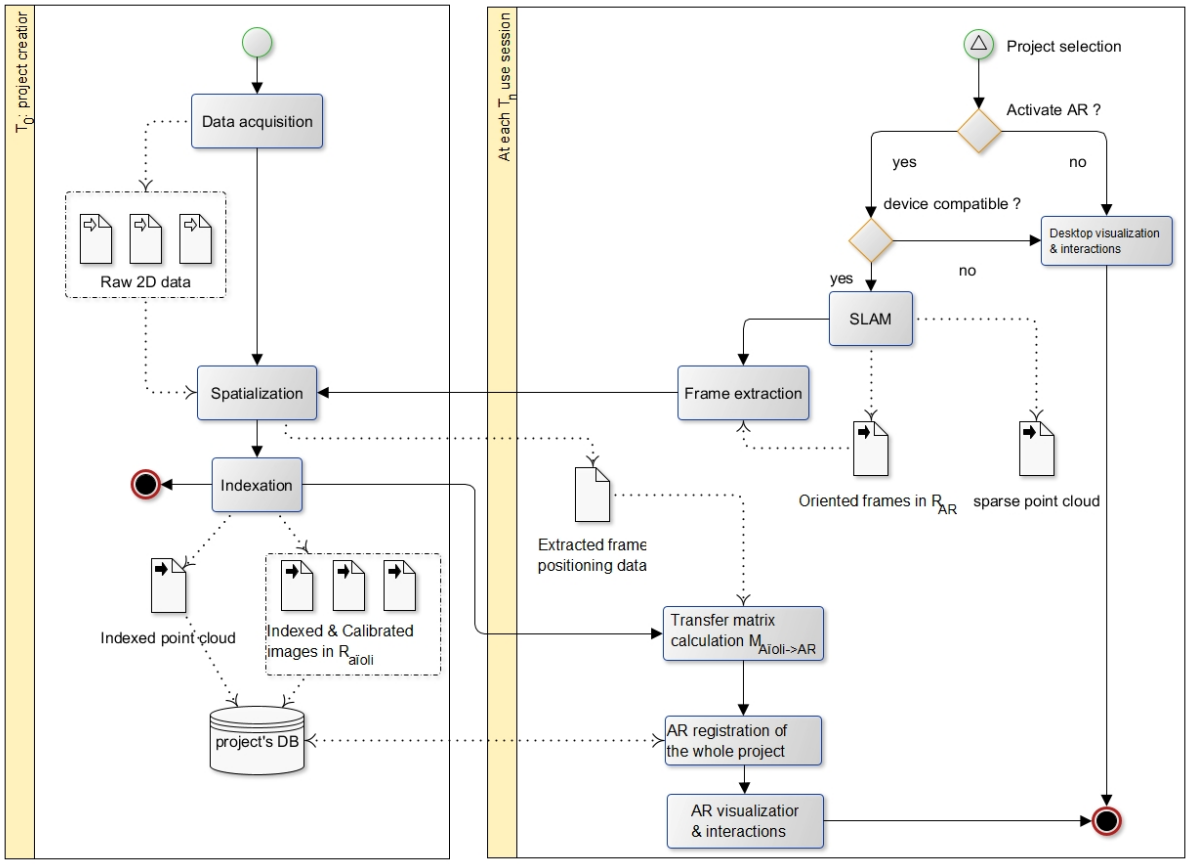

Figure 2: Overall approach for the registration of an Aïoli project. On the left, project creation at T_0. On the right, AR registration process at T_n, based on incremental photogrammetry.

project window. From this moment, through SLAM, we get in real time the estimates of the successive poses of the camera of the user device's video stream, and possibly a sparse point cloud according to the SDK used for the experiment [a]. If the user positions himself in front of his object of study and initiates the alignment, we can extract the active frame of the video stream and send it to the server for a spatial resection calculation, which will allow to estimate its pose in relation to the master images of the project. The pose of the same image, thus expressed in the two references systems, then allows to calculate the transfer matrix between the project reference system and the SLAMs one. We can therefore apply a 3D transformation to the whole scene to align it with the reality.

3. First step: a 2D/3D hybrid viewer

This scenario is based first and foremost on our ability to unify 2D and 3D content within a single representation space, whether for visualization or interaction. To do that, the first step is to literally translate the interrelations between 2D and 3D resources into a dynamic viewer by spatializing the different iconographic resources relative to a point cloud within a single scene. The aim is to create a coherent representative space manipulable in a single block. The implementation of this 2D/3D hybrid environment is an essential prerequisite, which we treated by developing a viewer from the ThreeJS (Cabello 2013) and PotreeJS (Schuetz 2016) WebGL libraries. It is structured as a 3D scene including a main camera and light sources, and in which we load the different entities of the project to be displayed: the point cloud (loaded as an octree (Schuetz 2014)), the oriented 2D photographs, and any annotations made by users.

During the spatial referencing process, the intrinsic and extrinsic parameters of each camera are written to a JSON file. This file allows us, at the viewer level, to generate equivalent virtual cameras, which are used to immerse the images in 3D space and navigate through all the acquisition dataset while interacting with the point cloud. Each photograph is visualized by applying it as a texture of a plane geometry, located on the near plane of the corresponding cameras frustum (Figure 3).

In terms of representation, images of the visible spectrum are often insufficient to visualize certain characteristics of the object of study, particularly in the case of rock art sites where environmental conditions and natural wall degradation can make observation conditions very difficult. Several other images are produced as a result of scientific analysis and contribute to the understanding of the studied object (Díaz-Guardamino et al. 2015; Domingo et al. 2015; Fischer and Kakoulli 2006; Rogerio-Candelera 2015). Nothing prevents these images from being displayed in the viewer, whether they are derived from calculations (normal maps, depth maps, roughness maps, decorrelated images, etc.) or sensors (RTI, IR, UV images, etc.) as long as their specifications are known or calculable. As such, this principle permits the correlation of analyses from different sources within the same space. All these images are associated with the different virtual cameras, either by being labelled "master" images (photographs used for the photogrammetric process) or "auxiliary" (additional views complementing an existing image).

4. Second step: real/virtual alignment from spatial resection calculation

To evaluate our process, we used an experimental browser developed by Mozilla™ for iOS (Mozilla Mobile 2018), whose purpose is to explore the possibilities of extending WebVR to AR/MR capabilities and thus to allow the creation of web-based augmented reality contents. This application is not a fully featured browser and is based on the ARKit SDK. It provides some capabilities regarding motion

[a] As this feature is not currently integrated into all SDKs made available through the WebVR or WebXR APIs, we have chosen not to rely on it for now.

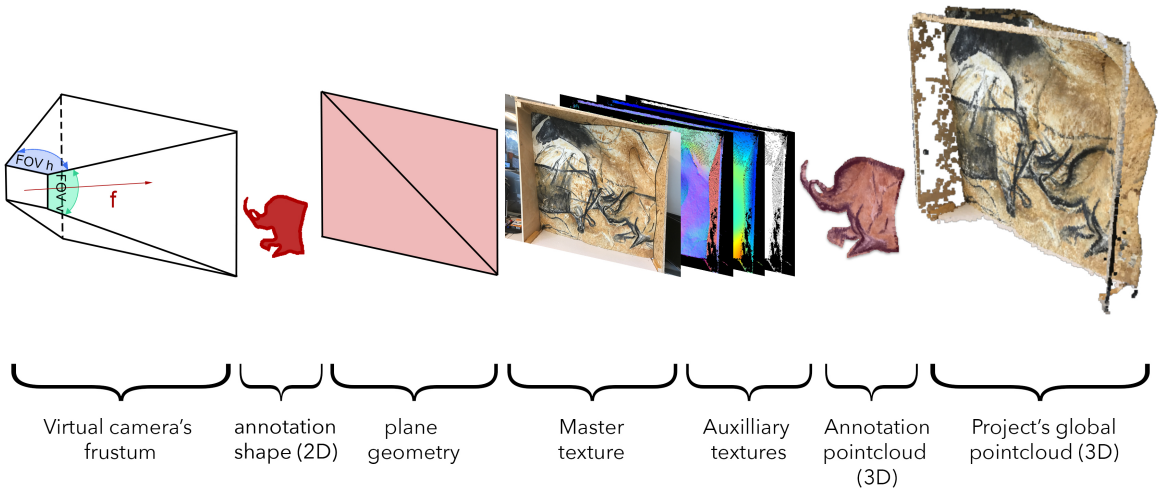

Figure 3: Structure of the 2D/3D hybrid viewer.

tracking, rendering of the pass-through camera, and some basic understanding of the real world like plane detection. To be noted that the GoogleAR team proposes a similar application for Android based on the ARCore SDK (Google-ar 2016). Besides, the WebVR specification is now becoming the WebXR Device API, and drafts are intended to gradually lead to standards.

For our very first experiment, we conduct tests on different datasets of six pictures of the same facsimile of the Chauvet-Pont dArc cave. Projects are created using these photographs, and annotations are made upstream in the desktop version of Aïoli. We then switch to the mobile version to turn on the AR mode.

In AR, all content of the project's 3D scene is rendered on top of the camera's video stream, and user controls are initialized to synchronize the movements of the camera with those of the device. At this point, the position of the scene is not coherent with reality. When the user starts the alignment, a frame is extracted, as a base 64 string and sent to the server for calibration. For this calculation, we can distinguish two cases: if the photograph to be oriented has useful metadata, the calculation can be based on the common Bundle Adjustment method (Deseilligny and Cléry 2011). Otherwise, the calculation must be based on the method of spatial resection (SR) by tie-points (11 parameters Direct Linear Transformation combined with RANSAC filtering) (Pamart, Morlet, and de Luca 2019).

Usually, the bundle adjustment method is considered as more precise, but also slower, which can be problematic in our case, partly for the user, who must wait until the end of the calibration to obtain the alignment of his project, and partly because SLAM can cause a progressive drift, which leads us rather to limit the calculation time to favour the time of effective use. In addition, as it stands, the camera flow is communicated frame by frame to the client in the form of a base 64 string and rendered in a canvas in the background. The images are then not provided with metadata, although we could add some on the fly, knowing the device used.

On the other hand, the spatial resection method is more versatile since it does not require metadata, and is also faster, but overall less accurate: if this calculation method provides a globally coherent orientation, the lack of metadata regarding the focal length can sometimes cause an irrelevant depth compared to the real position. Moreover, knowing that a 1-degree error represents a 17 centimeters deviation at a distance of 10 meters, it imposes to be vigilant because the quality of the AR alignment could be seriously compromised. We can therefore assume that the alignment method based on spatial resection will be more relevant for small scenes.

Assessing the strengths and weaknesses, we opted in a first step for the SR method due to its flexibility. Between the sending of the image to the server and the subsequent alignment, the computation time ranged from about 2 to 4 minutes depending on the situation. Several factors can influence this duration, mainly the definition of the new image, and its proximity to those of the initial acquisition set. Obviously, this calculation time is far too long to consider a comfortable use, but nevertheless reasonable for a first proof of concept. Moreover, even if the computation time is too long to consider a systematic use, the notion of anchor persistence allows to consider realistic operational scenarios, and interesting collaborative perspectives. By completing the alignment process by saving the space-mapping state and some key anchors from the world-tracking session, the alignment time for all subsequent AR sessions can be reduced to the loading time of the *WorldMap* containing all this information from the project database. In this way, our proposed registration procedure by spatial resection constitutes in fact a pre-calibration of the project that can benefit all the following sessions.

Visually, the alignment seems satisfying [b], since the annotations made on Aïoli are consistently positioned (Figure 4A). We can also see in transparency the proximity between the parietal figures of the facsimile and those of the point cloud (Figure 4B). A drift may however occur during the use of the application, particularly in the case of sudden movements by the user or in poor lighting conditions. Typically, this is characterized by a progressive shift of the scene from its initial position to a varying extent. During tests in a low-light environment, the phenomenon occurred several times, causing a drift of five to ten centimeters after several minutes of use.

Figure 4: Very first results of AR visualization of an Aïoli project with the Chauvet-Pont dArc facsimile. A: Dense point cloud. B: Sift cloud and dense annotation point cloud. C: Dense point cloud with normal texture. D: Composite texture (RGB and Normals).

IV. Planning the evaluation protocol

These early experiments are not yet sufficient to assess the accuracy of our method, nor its robustness to the various environmental variations that could affect rock art sites. A next step would be to find a protocol to evaluate our method. Ideally, the evaluation should be based on the comparison of the

[b] Additional videos can be found at this address: http://www.aioli.cloud/experiments/

point cloud from SLAM feature points and the projects one. As a starting point, the average distance and standard deviation could provide a first reliable indication of the quality of alignment. In addition, the use of the ICP algorithm (Besl and McKay 1992) to obtain the best matching between the two point clouds could also inform us about the proximity of our result to an optimal case. Finally, different approaches described in (Bogoslavskyi and Stachniss 2017) could be used to analyze the quality of matched 3D point clouds.

Nevertheless, several issues stand in the way of this perspective. The first one is the nature of the feature points extracted by ARKit. These points represent notable characteristics detected in the camera image during the session. As indicated by Apple© in the documentation, *"Their positions in 3D world coordinate space are extrapolated as part of the image analysis that ARKit performs in order to accurately track the device's position, orientation, and movement. Taken together, these points loosely correlate to the contours of real-world objects in view of the camera. ARKit does not guarantee that the number and arrangement of raw feature points will remain stable between software releases, or even between subsequent frames in the same session."* (Apple Inc © 2019). Indeed, feature points are not really depth points, and therefore do not provide a data stable enough to be used as a reference when comparing point clouds: such an evaluation protocol would not be reproducible. Secondly, due to the experimental aspect of the API, the use of WebXR reduces our ability to access ARKit data. In particular, the browser used for our experiment does not allow feature points to be collected on the client side, due to latency considerations. Even so, the coordinate system may change over time, and there would be no guarantee that the coordinates (x, y, z) obtained would be comparable from one frame to another.

In our case, the quality of the registration depends essentially on two factors: the accuracy of the spatial resection calculation on the one hand, and the accuracy of the camera poses obtained by SLAM on the other. Another approach would therefore be to evaluate these two aspects independently, in order to deduce a quality indicator for the overall method. However, this approach would only provide indirect and incomplete information.

Given the fact that the definition of an evaluation protocol in our context is not a trivial matter, we are currently trying to simplify the problem by reducing it to a 2D one. Our idea is to obtain a quality indicator based on the comparison of two images of the interface, taken both from the very same position, the tablet being fixed to a tripod. The reference image only represents a given frame of the video stream, i.e., the real object. The second image is the same, but additionally contains the 3D scene, with the point cloud positioned over the real object, following our registration process. These two images are compared using a common image matching algorithm, including the extraction and description of points of interest by SURF (Bay, Tuytelaars, and Van Gool 2006) and their matching with FLANN (Muja and Lowe 2009). The distance between each pair of matching points can then be calculated, as well as the average distance (Table 1). In the ideal case where the 3D model is perfectly registered, the expected result would be a near null average distance, as the two images would be almost identical. Similarly, the greater the registration error of the 3D model, the greater the expected average distance must be. Obviously, in practice the alignment is never perfect. For this reason, we obtained first a reference score by testing a manually positioned 3D object on its real position.

Point	Location on reference image	Location on right image	Distance (px)
Pt 0	(463.21, 572.193)	(459.873, 582.735)	11.0572
Pt 1	(760.528, 824.217)	(763.549, 839.645)	15.7211
Pt 2	(492.475, 706.755)	(490.371, 720.207)	13.6152
...			
Pt 859	(1138.92, 507.414)	(1150.45, 516.249)	14.5284
Pt 860	(1133.36, 654.276)	(1145.2, 664.3)	15.5131
Pt 861	(670.627, 562.544)	(672.916, 570.603)	8.37845

	References for this project	Our results
Average distance (px)	8.6912	13.6782
Median distance (px)	8.3322	15.2747
Standard deviation	4.8260	6.6885

Table 1: Example of the results obtained by our evaluation process. The average distance is compared with the reference distance, obtained with the point cloud manually positioned.

Although the early tests appear to confirm these expected results, we have not yet applied this method to a large enough number of samples to certify its effectiveness. In addition, it presents obvious limitations, since it prevents the assessment of the entire registration, which would require at least repeating this protocol on several points of view. It should still allow us to compare future results obtained by varying some contextual elements, such as brightness. This work is still ongoing.

V. Towards AR semantic annotation tools: some interaction perspectives

The project being properly positioned, we can already start exploring the possibilities in terms of visualization and interaction. The current structure already allows to visualize pre-existing annotations as 3D point clouds and 2D shapes. These annotations can be selected from the layers panel to display the related semantic and geometric descriptors. It is also possible to exploit the augmented reality mode to visualize certain geometric attributes directly in their real context, through the several visual descriptors calculated during the indexation step. The user only has to change the point cloud texture and switch from classic RGB colors to normal, curvature, roughness, index, elevation gradients, or even create a composite texture by playing on the opacity of these different information layers (Figure 4C-D).

We also explored different visual enhancement modalities, some of which are currently being implemented, in or-

der to improve users perceptual capabilities and help the analysis process. Decorrelation stretching, for example, is known to be a valuable tool for the study of rock art, which helps in the visualization of paintings, even discolored or degraded (Le Quellec, Duquesnoy, and Defrasne 2015). So far used to highlight some colorimetric characteristics on "classic" 2D photographs, we identified different ways to benefit from its capabilities in the scope of an augmented reality use. By applying this algorithm to a point cloud texture, we were able to complete the visual descriptor tool mentioned above, offering in this way to users a new annotation support more adapted to their observation environment (Figure 5). We also noticed interesting possibilities for observation support when applying this method to a real-time video stream, revealing color variations too faint to be visible to the naked eye.

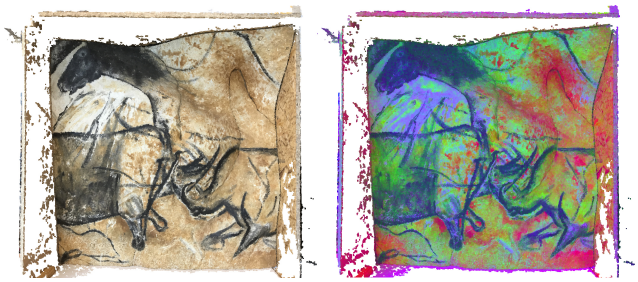

Figure 5: Projects point cloud with RGB texture (left) and decorrelated texture (right).

However, to be relevant for the study of rock art, such visualization modalities must be combined with tools ensuring the capture and storage of the observations made by experts. This is why we now aim to provide annotation tools in line with the great opportunities of augmented reality and the needs raised by the context of rock art. A magic wand annotation tool is currently being implemented and allows to automate the selection of 3D points of similar color to a reference value +/- a given threshold. Depending on the active texture, the magic wand can have different effects: applied to a normal texture, the tool is able to select the 3D points belonging to a same plane. Applied to a decorrelated texture, the tool allows to quickly select parietal figures. The existing raycast-based Aïoli 3D annotation tool is more generic and already allows regions to be drawn directly on the point cloud. Although the use of the different textures again facilitates the annotation process, this tool remains to be improved as it limits the annotable area to that covered by photogrammetry, regardless of the users potential wanderings around his object of study.

Finally, among the significant perspectives for the study and conservation of rock art sites are the possibilities for multi-temporal monitoring. Indeed, if the sparse cloud provided by SLAM does not seem sufficient to support relevant exploitation, the video stream frames associated with their positioning data could easily be subjected to the spatialization and indexation process at the end of each AR visit. In this way, users could access to data for comparing different time states as proposed by (Peteler et al. 2015) and thus ensure the morphological monitoring of wall degradations.

VI. Conclusion

In this paper, we described a method for aligning a reality-based scene in a coherent way on reality. We presented our first results, showing the possibility of visualizing 3D annotations made for the documentation of rock art sites. This constitutes a first step towards replacing the real object back at the heart of the survey process. Such an application seems to open many perspectives in terms of visualization and analysis support. Many modes of navigation and visualization are to be imagined to meet the challenges of our case study. In terms of traceability in particular, the backup of navigation data (user trajectory, supports and textures used for each annotation, etc.) could constitute a considerable improvement in the management of data, but also in the understanding of the relationships existing between the observer and the study object during a survey.

Although augmented reality seems to bring many opportunities, the acceptability of such an application by experts cannot yet be determined. For this reason, the ongoing work needs to be done in direct relation with the teams of experts involved in rock art analysis and conservation. The radical paradigm shift that these new tools would imply for them requires that issues concerning the ergonomics and acceptability of the proposed tools be raised in the very near future.

In addition, several process elements should be automated (i.e., scaling) or optimized. The calculation time for spatial resection, for example, is currently far too high and constraining. To speed up this process, we can on one side work on a better compromise between the definition of video frames and the global framerate of the application, and on the other side allow the manual pre-selection of the closest views to facilitate the later image matching step. We can also explore the notion of persistence, which is beginning to emerge in multiuser augmented reality experiments.

Finally, although our case study focuses for now on the documentation of rock art sites, this approach could well be adapted to other contexts of documentation, conservation or restoration of cultural heritage, whether buildings, sculptures, paintings, pieces of art or archeological fragments.

Acknowledgments

The authors would like to thank the *Centre National de la Préhistoire* and the *Syndicat Mixte de la Grotte Chauvet* for their support and for the loan of the facsimile.

References

[1] Apollonio, F. I.; Basilissi, V.; Callieri, M.; Dellepiane, M.; Gaiani, M.; Ponchio, F.; Rizzo, F.; Rubino, A. R.; Scopigno, R.; and Sobra, G. 2018. A 3d-centered information system for the documentation of a complex restoration intervention. *Journal of Cultural Heritage* 29:89–99.

[2] Apple Inc ©. 2019. rawFeaturePoints - ARFrame | Apple Developer Documentation.

[3] Arnaldi, B.; Guitton, P.; and Moreau, G. 2018. *Virtual reality and augmented reality: Myths and Realities*. Hoboken, NJ: ISTE Ltd/John Wiley and Sons Inc.

[4] Bay, H.; Tuytelaars, T.; and Van Gool, L. 2006. SURF: Speeded Up Robust Features. In Leonardis, A.; Bischof, H.; and Pinz, A., eds., *Computer Vision ECCV 2006*, volume 3951. Berlin, Heidelberg: Springer Berlin Heidelberg. 404–417.

[5] Bekele, M. K.; Pierdicca, R.; Frontoni, E.; Malinverni, E. S.; and Gain, J. 2018. A Survey of Augmented, Virtual, and Mixed Reality for Cultural Heritage. *Journal on Computing and Cultural Heritage (JOCCH)* 11(2):7.

[6] Bergstrom, L. 2018. New API to Bring Augmented Reality to the Web Mozilla Hacks - the Web developer blog.

[7] Besl, P., and McKay, N. D. 1992. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14(2):239–256.

[8] Blaško, M.; Cacciotti, R.; Křemen, P.; and Kouba, Z. 2012. Monument damage ontology. In *Progress in cultural heritage preservation: 4th International Conference*, 221–230. Limassol, Cyprus: Springer.

[9] Bogoslavskyi, I., and Stachniss, C. 2017. Analyzing the quality of matched 3d point clouds of objects. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 6685–6690. Vancouver, BC: IEEE.

[10] Boochs, F.; Trémeau, A.; Murphy, O.; Gerke, M.; Lerma, J.; Karmacharya, A.; and Karaszewski, M. 2014. Towards A Knowledge Model Bridging Technologies And Applications In Cultural Heritage Documentation. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* II-5:81–88.

[11] Buerli, M., and Misslinger, S. 2017. Introducing ARKit: Augmented Reality for iOS. Copyright © 2018 Apple Inc. All rights reserved.

[12] Cabello, R. 2013. three.js - Javascript 3d library. https://threejs.org/.

[13] Clini, P.; Frontoni, E.; Quattrini, R.; and Pierdicca, R. 2014. Augmented Reality Experience: From High-Resolution Acquisition to Real Time Augmented Contents. *Advances in Multimedia* 2014:1–9.

[14] Damala, A.; Hornecker, E.; van der Vaart, M.; van Dijk, D.; and Ruthven, I. 2016. The Loupe: Tangible Augmented Reality For Learning To Look At Ancient Greek Art. *Zenodo*.

[15] De Luca, L.; Pierrot-Deseilligny, M.; Manuel, A.; Chevrier, C.; Lollier, B.; Benistant, P.; Pamart, A.; Friederike, P.; Abergel, V.; and Alaoui, A. 2018. aïoli A reality-based 3d annotation cloud platform for the collaborative documentation of cultural heritage artefacts.

[16] Deseilligny, M. P., and Cléry, I. 2011. Apero, an open source bundle adjusment software for automatic calibration and orientation of set of images. In *Proceedings of the ISPRS Symposium, 3DARCH11*, volume 269277.

[17] Díaz-Guardamino, M.; García Sanjuán, L.; Wheatley, D.; and Rodríguez Zamora, V. 2015. RTI and the study of engraved rock art: A re-examination of the Iberian south-western stelae of Setefilla and Almadén de la Plata 2 (Seville, Spain). *Digital Applications in Archaeology and Cultural Heritage* 2(2-3):41–54.

[18] Doerr, M. 2009. Ontologies for Cultural Heritage. In Staab, S., and Studer, R., eds., *Handbook on Ontologies*. Berlin, Heidelberg: Springer Berlin Heidelberg. 463–486.

[19] Domingo, I.; Carrión, B.; Blanco, S.; and Lerma, J. 2015. Evaluating conventional and advanced visible image enhancement solutions to produce digital tracings at el Carche rock art shelter. *Digital Applications in Archaeology and Cultural Heritage* 2(2-3):79–88.

[20] Fischer, C., and Kakoulli, I. 2006. Multispectral and hyperspectral imaging technologies in conservation: current research and potential applications. *Studies in Conservation* 51(sup1):3–16.

[21] Fuentes, O. 2017. La 3d et l'étude de l'art pariétal, ses apports et ses limites. Technical report, Centre National de Préhistoire, Périgeux, France.

[22] Google-ar. 2016. WebARonARCore: An experimental Chromium modification in the form of an app for Android that lets developers build Augmented Reality (AR) experiences using web technologies on top of Google's ARCore. https://github.com/google-ar/WebARonARCore, original-date: 2017-08-29T14:24:42Z.

[23] Gosalia, A. 2018. Announcing ARCore 1.0 and new updates to Google Lens.

[24] Havemann, S.; Settgast, V.; Berndt, R.; Eide, Ø.; and Fellner, D. W. 2009. The arrigo showcase reloadedtowards a sustainable link between 3d and semantics. *Journal on Computing and Cultural Heritage* 2(1):1–13.

[25] Le Quellec, J.-L.; Duquesnoy, F.; and Defrasne, C. 2015. Digital image enhancement with DStretch textregistered : Is complexity always necessary for efficiency? *Digital Applications in Archaeology and Cultural Heritage* 2(2-3):55–67.

[26] Madsen, J. B., and Madsen, C. B. 2015. Handheld Visual Representation of a Castle Chapel Ruin. *Journal on Computing and Cultural Heritage* 9(1):1–18.

[27] Manferdini, A. M.; Remondino, F.; Baldissini, S.; Gaiani, M.; and Benedetti, B. 2008. 3d modeling and semantic classification of archaeological finds for management and visualization in 3d archaeological databases. In *Int. Conference on Virtual Systems and MultiMedia (VSMM)*, 221–228.

[28] Manuel, A.; Alaoui M'Darhri, A.; Abergel, V.; Rozar, F.; and De Luca, L. 2018. Quantitative Analysis and semi-automatic annotation of spatialized data. In *In press*. San Francisco, USA: IEEE.

[29] Marchand, E.; Uchiyama, H.; and Spindler, F. 2016. Pose Estimation for Augmented Reality: A Hands-On

Survey. *IEEE Transactions on Visualization and Computer Graphics* 22(12):2633–2651.

[30] Medley, J. 2018. Augmented reality for the web | Web.

[31] Mozilla Mobile. 2018. WebXR-iOS: An iOS app for testing WebXR. https://github.com/mozilla-mobile/webxr-ios, original-date: 2017-09-10T20:40:57Z.

[32] Muja, M., and Lowe, D. G. 2009. Fast Approximate Nearest Neighbors With Automatic Algorithm Configuration. In *Proceedings of the Fourth International Conference on Computer Vision Theory and Applications*, 331–340. Lisboa, Portugal: SciTePress - Science and and Technology Publications.

[33] Myers, D.; Dalgity, A.; Avramides, I.; and Wuthrich, D. 2012. Arches: An Open Source GIS for the Inventory and Management of Immovable Cultural Heritage. In Hutchison, D.; Kanade, T.; Kittler, J.; Kleinberg, J. M.; Mattern, F.; Mitchell, J. C.; Naor, M.; Nierstrasz, O.; Pandu Rangan, C.; Steffen, B.; Sudan, M.; Terzopoulos, D.; Tygar, D.; Vardi, M. Y.; Weikum, G.; Ioannides, M.; Fritsch, D.; Leissner, J.; Davies, R.; Remondino, F.; and Caffo, R., eds., *Progress in Cultural Heritage Preservation*, volume 7616. Berlin, Heidelberg: Springer Berlin Heidelberg. 817–824.

[34] Pamart, A.; Morlet, F.; and de Luca, L. 2019. A fully automated incremental photogrammetric processing dedicated for collaborative remote-computing workflow. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-2/W9:565–571.

[35] Peteler, F.; Gattet, E.; Bromblet, P.; Guillon, O.; Vallet, J.-M.; and De Luca, L. 2015. Analyzing the evolution of deterioration patterns: A first step of an image-based approach for comparing multitemporal data sets. In *2015 Digital Heritage*, 113–116. Granada, Spain: IEEE.

[36] Ridel, B.; Reuter, P.; Laviole, J.; Mellado, N.; Couture, N.; and Granier, X. 2014. The Revealing Flashlight: Interactive Spatial Augmented Reality for Detail Exploration of Cultural Heritage Artifacts. *Journal on Computing and Cultural Heritage* 7(2):1–18.

[37] Rogerio-Candelera, M. A. 2015. Digital image analysis based study, recording, and protection of painted rock art. Some Iberian experiences. *Digital Applications in Archaeology and Cultural Heritage* 2(2-3):68–78.

[38] Rolland, J. P.; Baillot, Y.; and Goon, A. A. 2001. A survey of tracking technology for virtual environments. In *Fundamentals of Wearable Computers and Augmented Reality*. Lawrence erlbaum associates edition. pp. 67–112.

[39] Rupnik, E.; Daakir, M.; and Pierrot Deseilligny, M. 2017. MicMac a free, open-source solution for photogrammetry. *Open Geospatial Data, Software and Standards* 2(1).

[40] Saint Aubin, J.-P. 1992. *Le relevé et la représentation de l'architecture*. Number 2 in Documents et méthodes. Paris: Inventaire général.

[41] Schuetz, M. 2014. PotreeConverter - Uniform Partitioning of Point Cloud Data into an Octree.

[42] Schuetz, M. 2016. Potree: Rendering Large Point Clouds in Web Browsers. Master's thesis, Institute of Computer Graphics and Algorithms, Vienna University of Technology, Vienna, Austria.

[43] Shi, W.; Kotoula, E.; Akoglu, K.; Yang, Y.; and Rushmeier, H. 2016. CHER-Ob: A Tool for Shared Analysis in Cultural Heritage.

[44] Soler, F.; Melero, F. J.; and Luzón, M. V. 2017. A complete 3d information system for cultural heritage documentation. *Journal of Cultural Heritage* 23:49–57.

[45] Taketomi, T.; Uchiyama, H.; and Ikeda, S. 2017. Visual SLAM algorithms: a survey from 2010 to 2016. *IPSJ Transactions on Computer Vision and Applications* 9(1).

[46] W3C. 2017. WebVR. Copyright ©W3C.

[47] W3C. 2018. WebXR Device API. Copyright ©W3C.

[48] Wefers, S.; Reich, T.; Tietz, B.; and Boochs, F. 2016. SIVTProcessing, viewing, and analysis of 3d scans of the porthole slab and slab b2 of Züschen I. In *Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods In Archaeology. Oxford: Archaeopress Publishing Ltd*, 1067–1080. Oxford, England: Archaeopress Archaeology.