


Article

A Robust and Versatile Pipeline for Automatic Photogrammetric-Based Registration of Multimodal Cultural Heritage Documentation [†]

Anthony Pamart ^{1,*}, François Morlet ², Livio De Luca ¹ and Philippe Veron ³ 

¹ Modèles et simulations pour l'Architecture et le Patrimoine, UMR 3495 CNRS/MC, 13402 Marseille, France; livio.deluca@map.cnrs.fr

² Centre Interdisciplinaire de Conservation et de Restauration du Patrimoine, LABCOM CICRP/MAP, 13003 Marseille, France; francois.morlet@map.cnrs.fr

³ LISPEN (EA 7515), Arts et Métiers Institute of Technology, 13617 Aix-en-Provence, France; Philippe.Veron@ensam.eu

* Correspondence: anthony.pamart@map.cnrs.fr

[†] This paper is an extended version of our paper published in *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XLII-2/W9, 2019 8th Intl. Workshop 3D-ARCH "3D Virtual Reconstruction and Visualization of Complex Architectures", Bergamo, Italy, 6–8 February 2019; pp. 565–571.

Received: 30 April 2020; Accepted: 21 June 2020; Published: 25 June 2020



Abstract: Imaging techniques and Image Based-Modeling (IBM) practices in the field of Cultural Heritage (CH) studies are nowadays no longer used as one-shot applications but as various and complex scenarios involving multiple modalities; sensors, scales, spectral bands and temporalities utilized by various experts. Current use of Structure from Motion and photogrammetric methods necessitates some improvements in iterative registration to ease the growing complexity in the management of the scientific imaging applied on heritage assets. In this context, the co-registration of photo-documentation among other imaging resources is a key step in order to move towards data fusion and collaborative semantic enrichment scenarios. This paper presents the recent development of a Totally Automated Co-registration and Orientation library (TACO) based on the interoperability of open-source solutions to conduct photogrammetric-based registration. The proposed methodology addresses and solves some gaps in term of robustness and versatility in the field of incremental and global orientation of image-sets dedicated to CH practices.

Keywords: close-range photogrammetry; Structure from Motion; MicMac; OpenMVG; data-fusion; remote-computing; multimodal imaging; Cultural Heritage

1. Introduction

Image-based modeling (IBM) states nowadays as one of the most used technique in the CH field thanks to its flexibility and cost efficiency compared to Range-based modeling. Nonetheless, it requires specific technical skill and expertise in data acquisition but also in the processing stage. Regarding this point, CH oriented applications are still split between the complexity of open-source packages and the opacity of commercial “black-boxes” to process the data. In this context the development of user friendly and open-source based solution seems a relevant choice for CH scientific purposes. However, combining and balancing the prerequisites between automation and accuracy are never an easy nor straightforward to goal. CH domain has also to deal with overgrowing massive and complex data while having to anticipate its over time management. Our contribution to this challenge is called TACO (Totally Automated Co-registration and Orientation of image sets). This toolbox formalizes

a methodological approach dedicated for photogrammetric documentation of CH artifact including features like iterative registration or interoperable functions. It has been developed initially as the photogrammetric engine for AIOLI, a reality-based 2D/3D annotation platform for the collaborative documentation of heritage artefacts (see Appendix A). Conceived as a flexible and evolutive tool it aims to support a global data fusion methodology allowing to handle and merge multiple sources of data. In CH the field, multimodality means to manage multidimensional data provided by the use of different sensors, scales, spectral bands and/or simultaneously acquired or in a multi-temporal context, to document and study heritage assets. Based on previous photogrammetric fusion experiments [1–3] detailing different combined modalities, this article presents recent advancements on automated co-registration of CH multimodal imaging. The development of TACO started and has been improved over the past few years on other joint research projects more especially focused on the documentation and follow-up of conservation-restoration works.

Multimodality in photogrammetric based workflows for CH documentation (see Figure 1) has involved some modifications from standard linear approaches, simplified hereafter as a synchronous Input and output (I/O) operation (Figure 1i). Nowadays, a presumed “final” output actually results from a complex operative chain involving scattered fusion and processing methods from multiple sources of data (Figure 1ii). During the co-registration of new data, we can face numerous issues for each step (shown in Figure 1iii); redundancy or compatibility with previous data (a), deprecated processing nodes (b), complexification and efficiency of fusion stage (c) or accumulation of disparate results (d). In this context, TACO aims to contribute to the creation of a stable and scalable method supporting incremental co-registration of CH resources over time (Figure 1iv).

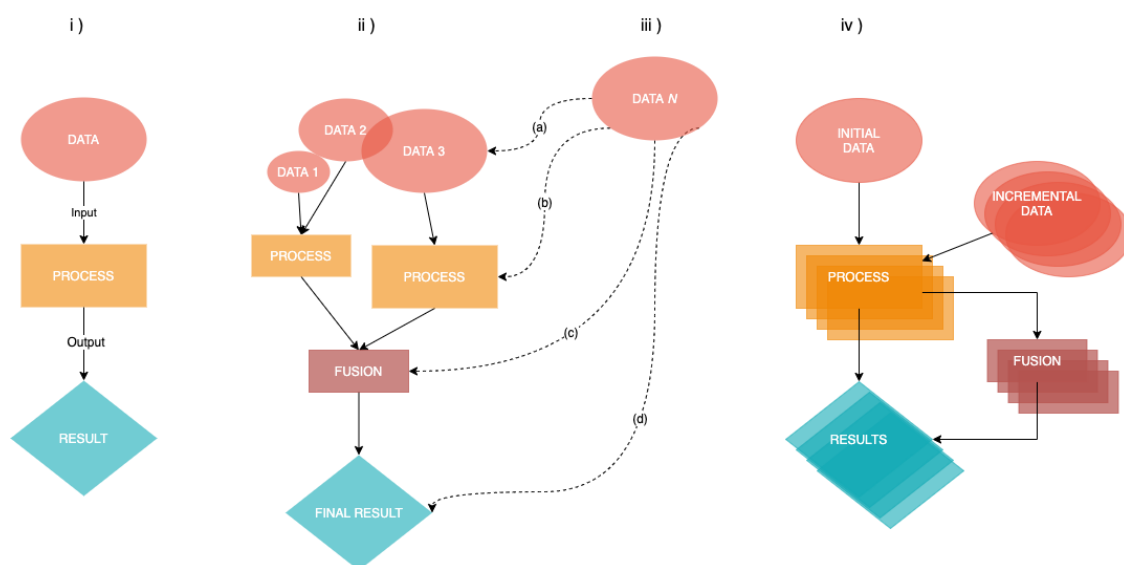


Figure 1. Diagram synthesizing CH oriented multimodal IBM workflows; conventional “one-shot” application (i), current state (ii) and issues (iii) and targeted approach (iv) using TACO.

1.1. Structure

The development of TACO has started on the assumption of the growing use of multimodal practices in the CH domain, introduced and defined above in Section 1. The next subsection interrogates TACO and related-works Section 1.2 in regard of IBM applications for data fusion purposes. Starting with a brief discussion on IBM framework, the Section 2 exposes all the components (in Section 2.1) and details the technical implementation for the *Initial* reconstruction (in Section 2.2) and *Incremental* registration (in Section 2.3) of multimodal image-sets. The results are presented and discussed in the Section 3, completed by our works in progress on interoperability problematic Section 3.1. Conclusion and perspectives on the methodological asset of TACO for CH data fusion are given in Section 4.

The Appendix A provides as an example, an operative implementation of TACO Appendix A within AIOLI web-service.

1.2. Related Literature

From the past decade, several domains such as the CH field target the digital replication as a mid-term scientific goal. This replica or “digital twin” of a CH artifact is facing two majors issues. The first one is the ability to build a digital replica as faithful and complete as possible and for which 3D scanning technologies provide great support. The second issue is the capacity of integrating seamlessly new data to keep the digital replica up to date according to the ever growing mass of digital documentation.

For the first issue, IBM is one of the techniques mainly used to generate a digital footprint of an object appearance at a given time. Affordable and considered as user-friendly, recent trends within both the scientific and not-expert community report on remarkable efforts to make it even more accessible and popular. The remote-computing applications (i.e., automatized and/or massive processing making accessible through a web-service) contributed to this goal since the early stage of digital photogrammetry [4–8]. Most of those works faced similar issues, i.e., the difficulty to reach a robust pipeline (meaning the ability to cope errors) and to define a versatile workflow, regardless of input data. Lately, the deployment of other web-services (e.g., WebODM, MiaW) have been eased with the automation degree included in FLOSS packages [9,10]. The development of photogrammetric Software as a Service (SaaS) has emerged concurrently with the upgrade of SfM solutions for massive processing. After decades of parallel development, the offer in IBM software remains fuzzy between SfM algorithms coming from Computer Vision community and more conventional photogrammetric methods [11,12]. As stated recent work [13], despite their similarities and differences both methods seem to converge toward an optimal and hybrid solution. Indeed, this so called “photogrammetric computer vision approach” might be a relevant solution to address the challenge of massive digitization of CH. On one side SfM has proven since the beginning its robustness and velocity to handle massive and unstructured dataset [14]. On the other side, only the controlled environment of photogrammetric approaches can perform a global refinement [15] with demanding camera self-calibration models [16] to reach accurate measurements required for CH multi-source 3D dense reconstruction. In order to fully benefit of the assets belonging to each method, library or algorithm, the need of an interoperable process is raising. Some attempts have been recently made in this direction to release more handy tools [17,18]. Furthermore, the CH experts have high demand on preserving a data-provenance and data-quality continuum, hence the use of commercial solutions is frequently questioned in cross-comparison studies [19–21] in term of cost to performance ratio.

The second issue concerns the spatial registration of images which is the fundamental step prior to address the challenge of Data Fusion in the CH field. Several related works achieved to link or bridge imaging modalities by exploiting feature-based registration method combined with photogrammetric routines, in multi-band [22–24], multi-source [25–27] and/or multi-scale [28] approaches. As stated in a review article [29], the diversity of CH objects increases the complexity of defining and automating efficient or reproducible methods to reach satisfying data integration for CH purposes. One explanation of this complexity is the discrepancy of 2D and 3D data sources in the co-registration process compared to other domain like medical or satellite imaging where intermodality is better known and defined. For this purpose, TACO has been limited on image-based methods to avoid such technological gap and consequently improve the performances with photographic-based techniques widely used in CH documentation (e.g., photogrammetry, technical photography and computational photography). However, the extension of the co-registration scope to other techniques is planned and discussed in section Section 4.

This literature review underlines the current difficulty to build and handle the co-registration from several multimodal layers within a single and reproducible framework. This situation occurs

because of the richness and the variety of heritage assets themselves as much as for the diversity of analysis methods required for their scientific documentations.

2. Totally Automated Co-Registration and Orientations

IBM is stated in the literature as a stable and reliable technique despite its relative uncertainty [19]. As it remains an indirect measurement technique, the quality of the output is essentially related to the quality of the input data. In conventional applications the data processing is carried out by the same expert or institution, in charge of the adequacy and the control of data quality among the complete operative chain. The remote-computing and/or collaborative scenarios introduce a change of paradigm in that the data characteristics (i.e., size and/or complexity, quality, acquisition strategy and purposes of the survey) are variable and mostly unknown. Moreover in CH applications sub-optimal circumstances in the acquisition stage are numerous: on-site conditions, bad lighting, object shape or material, lack of end-user experience. The occurrence of one or more of those elements proportionally increases complications in further processing stages. Therefore considering the unknown complexity of the input data it becomes harder to define a robust processing (i.e., minimizing bias, errors and failures) which requires to compromise between balance between versatility, velocity, accuracy and automation. Our approach is initialized with parameters prioritizing on the accuracy and progressively release constraints to gain in robustness against a sub-optimal condition. The counterpart is the potential increase of uncertainty—still better than a failure—and computational time (compensated with our computation server capacities).

2.1. Overview and Technical Description

The photogrammetric routine is composed of two main modules processing as follows: given I_0 the *Initial* set of images we want to find the set M_0 of point correspondences between overlapping images, the set of (intrinsic and extrinsic) calibrations C_0 , and the set D_0 of dense 3D points in the scene. For the *Incremental* processing, with I_n the new set of images, we compute point correspondences M_n between images in $I_0 \cup \dots \cup I_n$, and then new calibrations C_n . Considering this routine as a linear succession of I/O processes TACO introduces loops in the decisive steps and tries to resolve errors by incrementing some key parameters. The first node operates a robust *Initial* reconstruction described in Section 2.2 and $Iter_0$ of Figure 2 with the ability to handle in one shot massive and complex image sets (including multiple sensors, focal). The second module allows a versatile *Incremental* co-registration to merge new image sets with previous iterations. The latter is divided into sub-modules deriving to different methods implemented and detailed in Section 2.3 and $Iter_n$ of Figure 2. Optional modules are articulated to manage additional features like: absolute orientation based on coded-targets, interface to AIOLI or interoperable layers. TACO functions can be combined to build a fully automated pipeline suitable for remote-computing scenario. However modules and sub-modules can also be parameterized and used separately. As an example, options have been implemented to skip sub-steps of a process (e.g., avoid to recompute 2D image matching or disable 3D dense matching) so as to improve interactive uses. This modulative conception offers the flexibility to handle variable multimodal dataset requirements, as the framework dedicated to RTI fusion presented in a previous work [3].

The main dependence used as the core engine for photogrammetric reconstruction rely on MicMac, developed mainly at IGN since 2003 [10]. Also known as the *Pastis-Apero-Micmac* toolchain it has the benefit of having been evaluated in a metrological context [30] contrary to FLOSS alternatives. However like all solution it suffers from bottlenecks such as latencies in 2D matching (*Pastis*) or the lack of parallelized bundle adjustment (*Apero*). Lately, interesting performances in features extraction and matching steps have been achieved (see Figures 9 and 10) with OpenMVG [9]. At short term, TACO aims to combine those two packages efficiently before looking forward the integration of new libraries in close future (COLMAP, THEIA, AliceVision). Our on-going developments on interoperable features between OpenMVG and TACO are discussed in Section 3.1. TACO has been coded in a generic iterative/recursive script running through a Python program to call commands available in IBM

packages. An example of complete pipeline combining *Initial* and *Incremental* processes is structured as follows:

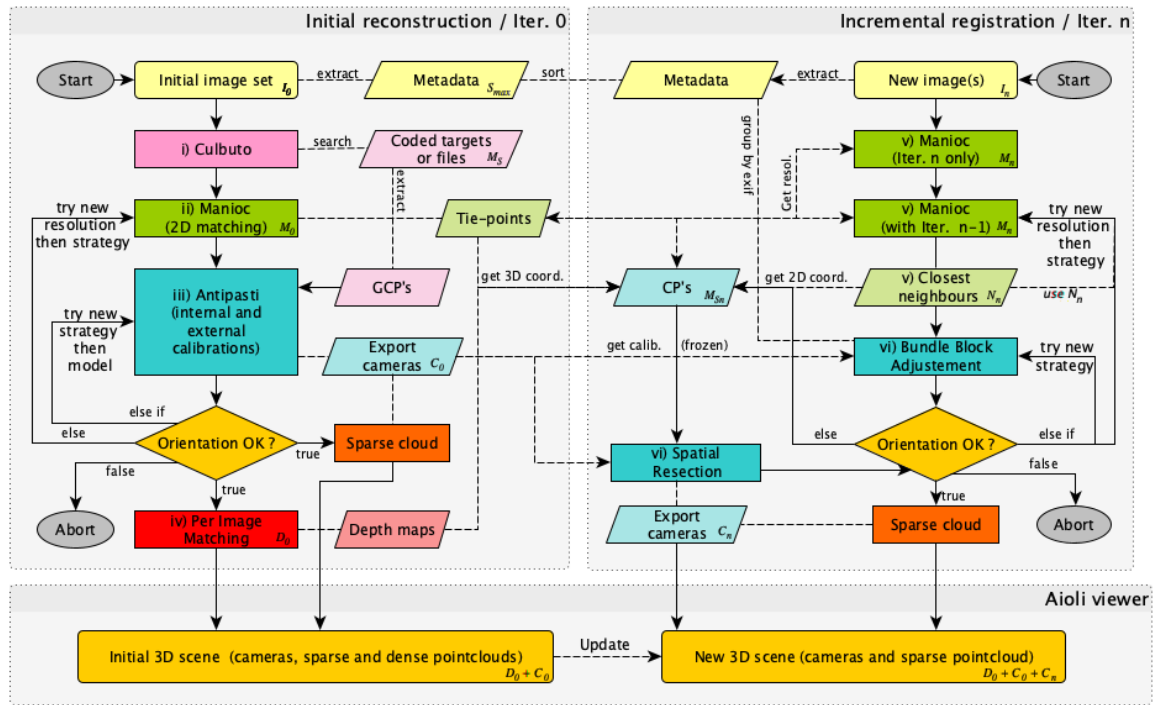


Figure 2. Diagram showing the example of the fully automated TACO-based pipeline within AIOLI.

- (i) *Culbuto* is the module for automatic scaling and georeferencing (see Figures 3 and 4): it takes initial image set I_0 and outputs scaled image correspondences M_S .
- (ii) *Manioc* is an automatic sub-sampling parametrization for 2D matching: it takes I_0 , and s_{max} the largest dimension of the images in I_0 , subsampled at scale $s \in \{1000px \dots s_{max}\}$, and outputs image correspondences M_0 .
- (iii) *Antipasti* is an adaptative self-calibration method: it takes M_0 , M_S and outputs external and internal calibrations C_0 .
- (iv) The fully automatic *Per Image Matching* tools produces the dense cloud D_0 from I_0 , M_0 , C_0 .
- (v) During incremental registration, *Manioc* takes I_0 , I_n new image set and produces M_n (correspondences with all image sets). It also gives the closest neighbours of I_n : $N_n \subset I_0 \dots I_{n-1}$.
- (vi) Finally, a robust and versatile incremental co-registration is performed given C_{n-1} , M_n , N_n , TACO computes C_n , in the same spatial system as C_0 , either by *Bundle Block Adjustment* or *Spatial Resection*.

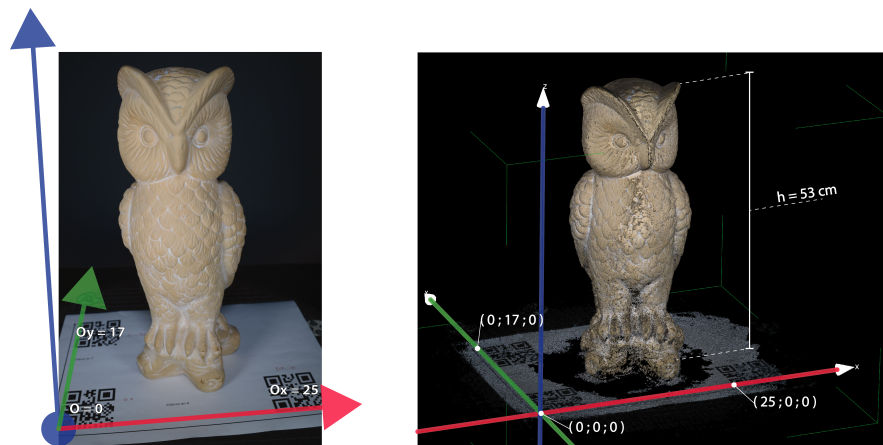


Figure 3. Owl dataset : example of QR-code triplet based global orientation and scaling.



Figure 4. Example of automated scaling and orientation of wood painting base on experimental double sided coded-targets fusion.

2.2. A Robust and Automated Initial Reconstruction for Complex Data Sets

The goal of this *Initial* process is to enhance the robustness, i.e., the ratio of successful computation, notwithstanding the type, the size, the complexity and the quality of the image set given in input. The *Initial* module passes through typical routine: 2D image matching from features extraction, internal self-calibration to relative orientation and absolute orientation until the sparse and dense cloud reconstruction. Beforehand, image meta-data are checked essentially to avoid missing EXIF (e.g., focal length) error or bias (e.g., image autorotation). A synthetic log is provided containing all meaningful camera settings for photogrammetry purposes (e.g., ISO, aperture, white balance etc). This approach is intended to guide the (re)processing strategy by EXIF grouping. Later on, it will facilitate the management, navigation and retrieval large oriented image-sets within AIOLI 2D/3D viewer (see Appendix A). Finally, this metadata aware process is requirement to strengthen the informative continuum in respect to data provenance scheme (semantic enrichment, CIDOC CRM etc.)

In (i), a module called *Culbutto* searches for 2D/3D coordinates files or tries to detect it from coded-targets within the images to compute absolute orientation in (iii). A first approach, is based on single QR-Code detection which embeds a metric dimension and where axes are derived from code corners (as supposed planar). It allows to approximate consistent scaling and orientation according to a user-defined coordinates system. This uncomplicated approach can be improved by using QR-code triplet as shown in Figure 3 devised for bigger object or scene. This method similar as using calibrated photogrammetric scale-bars still suffers from a lack of versatility and scalability for complex and/or

large scale acquisitions. Lately, a second approach has been experimented to integrate free-from GCP's networks; by providing conventional space and images files coordinates, or in an automated way by coupling them with coded-targets sequential detection. With this method, the indexes contained in the targets are matched with 3D measurements (acquired by telemetric method, laser scanner similarly to total station). Interesting experiences using CCTag [31] and AprilTag [32] have been lead to wield TACO in advanced multimodal data fusion scenario, see Figure 4.

In (ii), an iterative multi-resolution 2D features extraction and matching approach has been implemented to find an optimum budget of key-points aiming to optimize further internal and external calibration steps. Based on the MicMac tool *Tapioca*, it is starting with a low pass intended to speed-up the computation and the sub-sampling is recursively increased to medium and high passes (respectively half and three quarters of the full size) until camera poses are determined successfully. As shown in Table 1, the average Reprojection Error (RE) is usually around 1 pixel. Optionally, some advanced options enable to modulate the image matching ratio [33], to enhance local features [34] in order to cope rebellious cases (e.g., texture less or low-overlapping). Exploiting this strategy called *Manioc*, TACO helps to converge to a successful bundle solution in the next calibration steps.

In (iii), the relative orientation is achieved by a combined use of MicMac tools to compute the initial value of orientation (*Martini*) and refine internal parameters and global orientation (*Tapas*) for each camera. Those tools are respectively using triplets with Tomasi-Kanabe method while the other solves a more conventional Bundle Block Adjustment with the Levenberg-Marquardt method. The main difference is that *Tapas* is mandatory as the only one computing internal self-calibration. TACO's contribution is to alternate logically between those complementary tools until the condition of global orientation of all pictures is satisfied. In case of failure the calibration model is progressively decreased (i.e., from Fraser Basic to Radial Basic), else a return to remaining image matching strategies is attempted. By the mean of this processing methodology, called *AntiPasti*, the chance to return a reliable (minimized residue) and consistent solution for cameras calibrations have significantly increased.

Finally in (iv) a 3D dense matching algorithm computes a dense point cloud. It is performed by the core algorithm of MicMac, but TACO gives access to all *Per Image Matching* automated tool to adapt strategies and density parameters (using a pyramidal approach). The gain of using TACO is related to the improvement mentioned which affect positively on the uncertainty of the reconstruction. The possibilities to use other algorithms will depends on interoperable bridges discussed below in Section 3.1.

Therefore, this first iteration of the pipeline constructs a stable and reliable initial geometric reference (see Table 1) pending for new images to be added during the next iterations. This dataset is serving as base for the following co-registration thus it is assumed to be controlled and complete as defining a *master acquisition*. It means that the camera positions and resulting point cloud are (at this stage of development) frozen (initially because of AIOLI requirement detailed in Appendix A). Nevertheless a more flexible methodology to release and or freeze constraints on camera calibration subsets is planned for future works. Indeed this global refinement strategy involved correlated metrics and has to be conducted jointly to the qualitative assessment of 3D dense reconstruction in multiple sources context.

Table 1. Composition and results of initial data sets extracted from the global benchmarking.

Dataset	Initial nb. of Images	Computational Time	Sparse Cloud (k)Points	Dense Cloud (M)Points	Initial Avg. RE in (px)
SA13	8	2 min	102	2.35	1.08
Fragment *	21	4 min	60	1.34	0.47
Saint-Trophisme	26	7 min	94	3.13	0.54
Owl	40	7 min	108	2.68	0.93
Vasarely	65	7 min	127	1.76	0.74
Excavation site *	95	43 min	641	1.34	0.49
Small temple *	134	120 min	1850	4.04	0.74
Old-charity	181	170 min	224	1.14	0.73
Arbre-serpents	273	390 min	1468	5.62	1.26
Nettuno	256	Nc.	2806	9.87	0.43
Lys model	841	Nc.	20,989	44.4	0.66

(*) refers to unillustrated dataset for copyright reasons but their names reveal their typologies.

2.3. A Versatile and Incremental Spatial Registration of New Images from Oriented Image Set

The collaborative and multidisciplinary framework of AIOLI platform detailed in Appendix A, has originally set the incremental registration of images as a priority. Furthermore, the specificity of CH studies requires to face the challenging variety of multimodal scenarios. The main difficulty to deal with, is the unknown overlapping ratio between iterations while images might have an high deviation in spatial or pixel resolution, radiometric consistency and changes within the scene or the object itself.

Just as for the *Initial* reconstruction, the enhancement of features extraction and matching approaches (i.e., strategy and resolution) in (v) has revealed as an essential steps to succeed the co-registration. In the case of single picture, a matching based on heuristic approach is trivial and time-consuming especially if the dataset exceeds hundred of images [35]. In this context, our solution in consist of finding correspondences in a very low resolution to detect adjacent pictures to use as best-matches. In case of a new block of pictures, they are matched together before being matched with all or a subset of images from previous iterations. The *Manioc* strategy is repeated with the difference that the incremental sub-sampling benefit from past matches and a best-neighbours selection is used to speed-up computation. A lot of feature-based optimizations can be achieved at this stage. So far, some key-points filtering approaches have been experimented upon the findings of recent works [36].

For the spatial registration in (vi), two alternative methods are integrated as sub-modules. They are both initialized from the results of the previous iterations (i.e., key-points, internal and external calibration or the dense point cloud). As formerly explained, the previous camera poses and corresponding geometries remain fixed for the ease development, but a more flexible global refinement module is planned. A first *Bundle Block Adjustment* (BBA) based approach is dedicated if a regular photogrammetric dataset—understood a sequence of overlapping images—is given in input. However according to the differences in the resolution (image size, focal length or distance to the object), radiometry (exposure, spectral band, illumination) or context (parasites, alterations) BBA could converge to sub-optimal, biased or erratic solution because of a lack of key-points (usually below 1000 matched features). In this case, a surrogated approach is based on a *Spatial Resection* approach [37] to calculate a single camera pose (or several processed sequentially), which also has the benefit of being robust to missing meta-data (non-photographic sensor, archive, analog photography). This alternative is based on 11 parameters Direct Linear Transformation (DLT) combined with RANSAC filtering (for outliers removal) to estimate both internal and external calibrations among a known set of camera poses. As no Ground Control Points are available, we use the key-points as 2D coordinates and their projection onto the depth-maps to retrieve corresponding 3D coordinates. This SR method can return an approximative solution, but enables to find a homographic relation with only tens of matching points when BBA is impossible.

For BBA and SR approach as well, the global accuracy and correctness is not related only to the number of key-points used for poses estimation but also in their redundancy and spatial distribution [36]. As a result, lower is key-points abundance and multiplicity, higher will be the uncertainty of the calibration and orientation computed with SR. Anyhow with both methods a result can converge to a false or odd solving for which some refinement methods will be investigate. For troubleshooting or optimization purposes different leads conduct to the incremental bundle adjustment of OpenMVG [9,21], the transition to an up-to-date perspective-three-point solver [38–40] or trying to refine registration with 2D/3D Mutual-Information [41].

Despite some possible improvements, current results (see Table 2) seem promising in regard of forthcoming evolution of TACO to endorse its aptitude to process complex multimodal CH imaging data sets. Overall results in term of quality, relative accuracy and versatility are discussed and illustrated in the next section.

Table 2. Details of multimodal dataset enrichment and results of incremental processing.

Dataset	Added Images	Number of Iterations	Number of Sensors	Number of Focal Length	Final Avg. RE in(px)
SA13	+107	12	1	6	1.54
Saint-Trophisme	+4	4	4	4	0.6 *
Vasarely	+6	3	1	2	0.80
Old-charity	+1	1	1	1	1.12 *
Arbre-serpents	+259	3	1	1	1.65
Nettuno	Na	1	2	2	0.59
Lys model	Na	1	2	3	1.21

(*) refers to values including images oriented with Spatial Resection.

3. Results and Discussions

Instead of using synthetic data, close to a hundred of datasets from MAP laboratory archives have been reprocessed to test the robustness and the versatility of TACO. This benchmark was composed of variated projects; from 5 up to 300 images, acquired different sensor sizes and resolutions (from smart-phones to full-frame DSLR) with a wide range of focal length (from 12 mm to 105 mm). TACO successfully completed 99% of those image sets made up of various qualities, scales, complexities and objects reflecting the reality of CH application. The main condition of success is the respect of basic rules of image capture. Since it has been integrated to the latest release of AIOLI, beta-testers feedbacks globally noticed less failing process or issues related to photogrammetric stages. An extract of pre-selected CH object typologies from this massive benchmark is given in Table 1 for the initial iteration. The results of incremental processing, depicting the following multimodal scenarios, are given in Table 2:

- The *Arbre-Serpents* dataset (Figure 5) demonstrates the versatility to manage complex UAV-based data-acquisition.
- The *SA13* dataset (Figure 6) proves the ability to handle varied multi-focal acquisition.
- The *Vasarely* dataset (Figure 7) highlight the capacity to deal with multi-scalar acquisition (14× GSD magnification).
- The *Old-Charity* dataset (Figure 8) illustrate the velocity (40s of computing) to register an isolated picture on large image-set.
- The *Saint-Trophisme* dataset (Figure A2) shows the robustness to integrate multi-temporal images including web-archive.

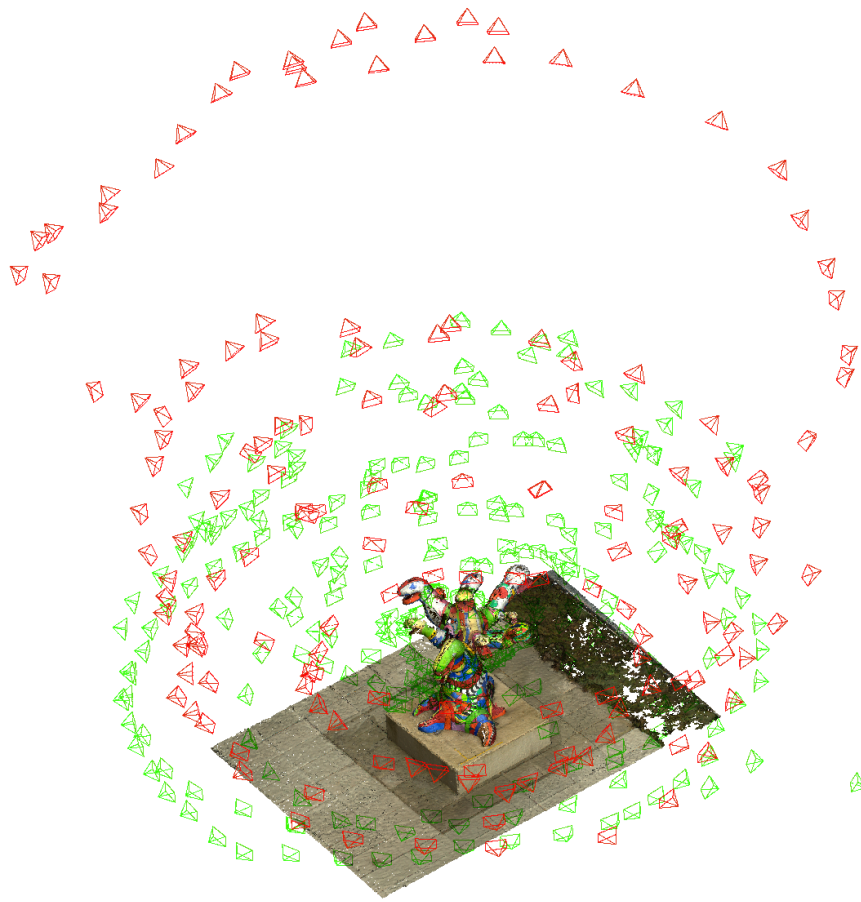


Figure 5. *Arbre-Serpents* dataset : fusion result of two UAV-based blocks of acquisition (before and after restoration).

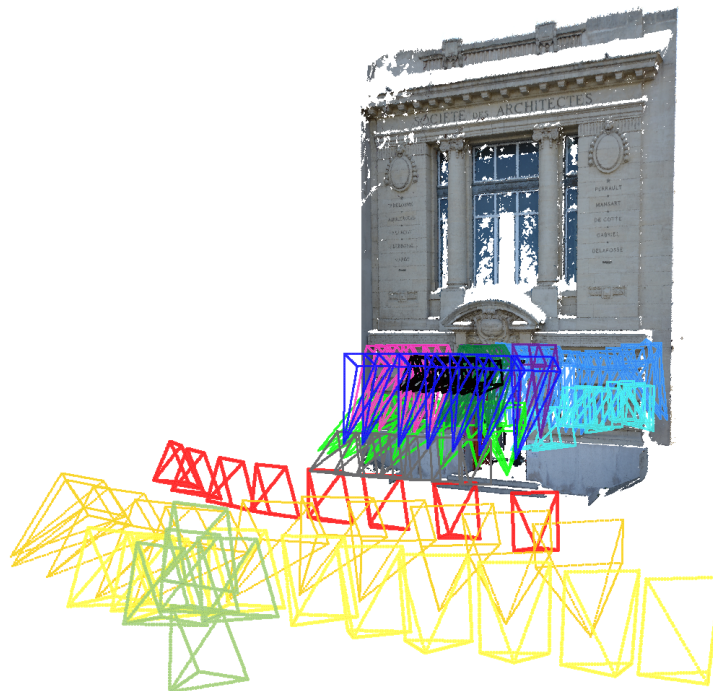


Figure 6. *SA13* dataset : example of complex multi-focal processing (each color represents a new iteration).

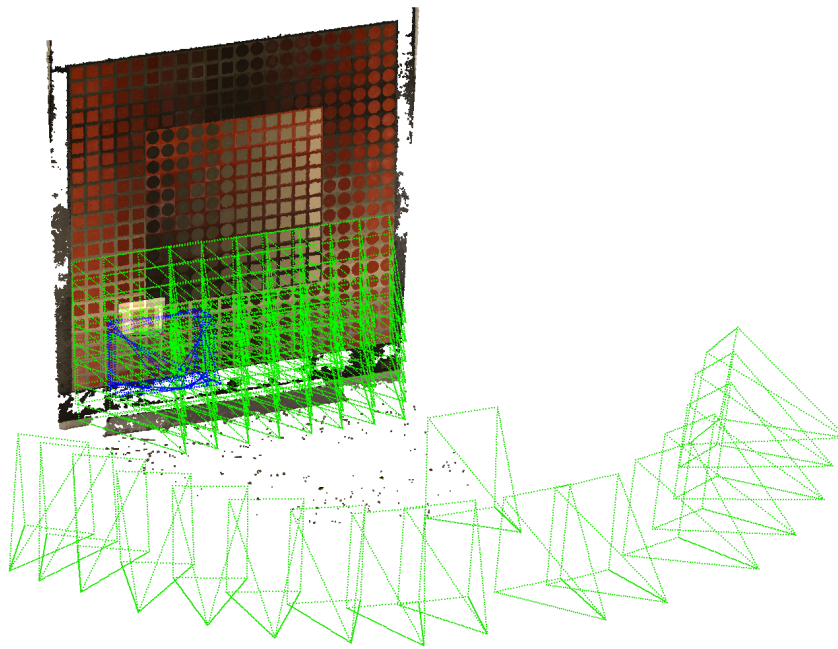


Figure 7. *Vasarely* dataset : data fusion with high scaling deviation (*Initial* iteration in green and *Incremental* set in blue).

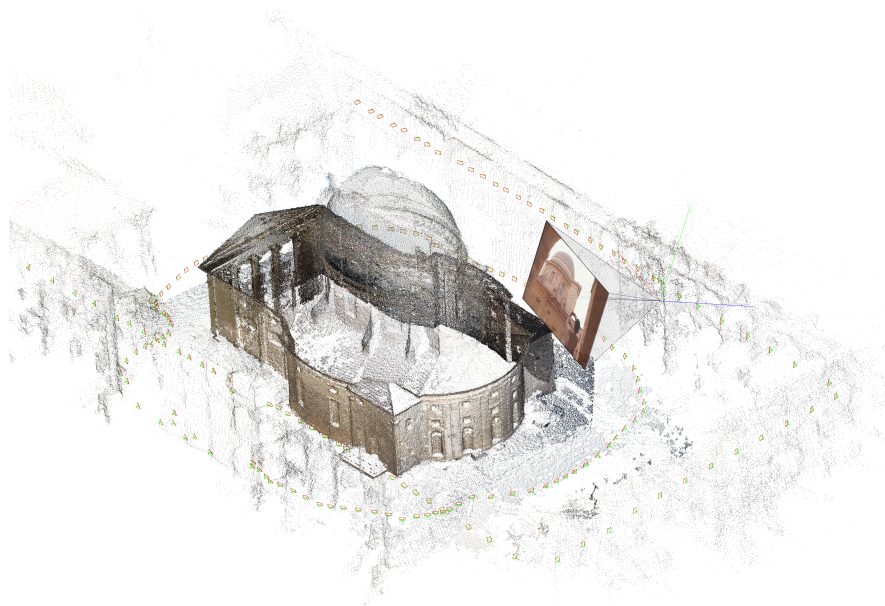


Figure 8. *Old-Charity* dataset : isolated single picture registration using Spatial Resection.

The benchmark we present here have been processed using the computation server devised for AIOLI (Dell Power Edge R940, with a Intel Xeon CPU for a total of 144 cores of 3.0 GHz, 256 GB of RAM and 4 TB SSD storage). A virtual environment with Docker container [42] is used to run processes on the server. Therefore, computational times given (in Table 1) may vary in case of server load or using a local workstation. The metrics for evaluating our benchmarking were the numbers oriented images and their theoretical precision in positioning which is assessed by an encouraging RMS reprojection error around 1 pixel. The necessary quality check and robustness evaluation will be verified in a further works by integrating a ground truth data comparisons to assess TACO performances.

3.1. Toward the Interoperability Challenge

In spite of fully automated pipeline advantage, this solution isn't exempt from drawback. Consequently we tried to compensate the lack of interaction in both *Initial* (see Section 2.2) and *Incremental* (see Section 2.3) modes by the possibility to modify somehow the linearity of the algorithm (modules, submodules, skip/break point or different tweak and options) to expand its potential. This modular architecture of TACO was also in anticipation of our future work in terms of interoperability.

Data and information exchange in IBM is a great and open challenge arising. Currently the several conventions coexist in the SfM and photogrammetric softwares while reliable exchange file formats remain restricted. Therefore the definition of interoperable functions is a key notion to improve any IBM operative chain in term of velocity, robustness and relative accuracy by exploiting the best features of distinct solutions to reach an optimum result. It can be used to seize the opportunity to bridge photogrammetric community and SfM approach, and inversely. Nowadays there is no straightforward solution to combine a fast and robust initial poses estimation on a massive and unstructured image set globally refined with a precise compensation on GCP's with demanding internal calibration model. Finally, improving interoperative dialogue between key processes is also a leverage toward open-data and open-science as it can offers a viable and accessible alternative to commercial solutions but also creating a reciprocal data flows from FLOSS or to FLOSS reprocessing. Recently, attempts have been made to interlink TACO with Agisoft Metashape for complementary purposes rather than comparative reasons. Our work in progress focusing on the interoperability between algorithms, libraries and commercial software solutions revealed the importance to solve numerous exchange and conversion issues enumerated below:

- A reversible tie-points format between single and multi-image observations.
- A lossless intrinsic calibration model according to variable intrinsic parameters conventions.
- A generic format for external calibration preserving uncertainty metrics.
- A human/machine understandable format to manage 2D/3D coordinates of GCPs.
- An enriched pointcloud format to compare and evaluate multi-source dense matching results.

For this purpose, two additional datasets are given as an example of the upcoming data conversion features offered by TACO. Those figures illustrate the benefit of a co-processing shared between OpenMVG and MicMac. The experimented pipeline (currently under development and implementation within TACO) uses OpenMVG until coarse initial orientation then data are converted and transferred to MicMac which performs a global refinement of intrinsic and extrinsic parameters and compute a dense cloud.

- On *Nettuno* dataset (Figure 9) the interoperable process solved critical 2D matching issue between very low overlapping terrestrial and UAV acquisition.
- On *Lys model* dataset (Figure 10) through the interoperable process a complex dataset composed of 841 was oriented with a single automated iteration.

The first example of *Nettuno* shows the limits of an approach based on a single solution. For this dataset, the MicMac-based attempts were failing among several iterations and optimisation trials while switching to OpenMVG unlock the situation. On the second example *Lys model* OpenMVG strongly speeded-up processing time while helping MicMac to reach trustworthy reprojection error (see avg. RE in Table 2) with a better camera model (Fraser Basic instead of Radial Basic). Therefore, the dense point cloud obtained was less noisy than the one processed straight from OpenMVG. Ideally, interoperability aim to achieve seamless conversion between all sub-processes to enrich and optimize operational chain leading upward to the highest output quality.

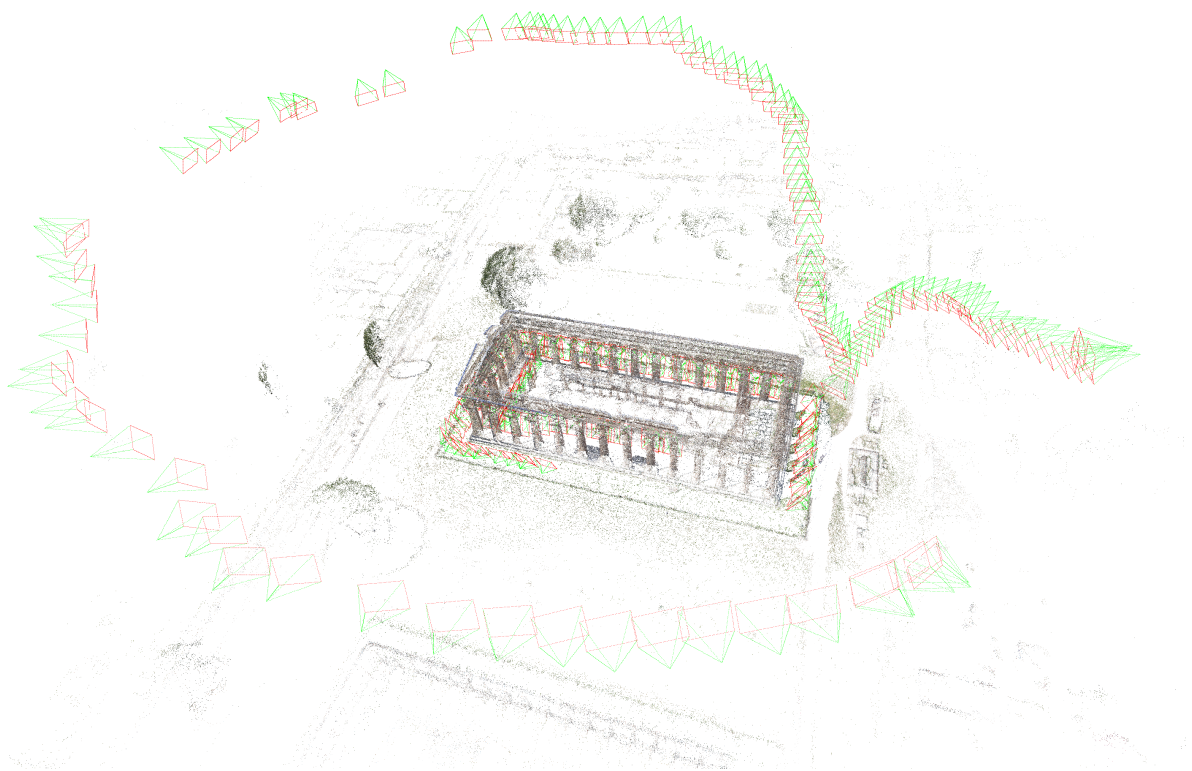


Figure 9. *Nettuno* dataset : successful interoperable OpenMVG+MicMac process on low overlapping UAV and terrestrial.

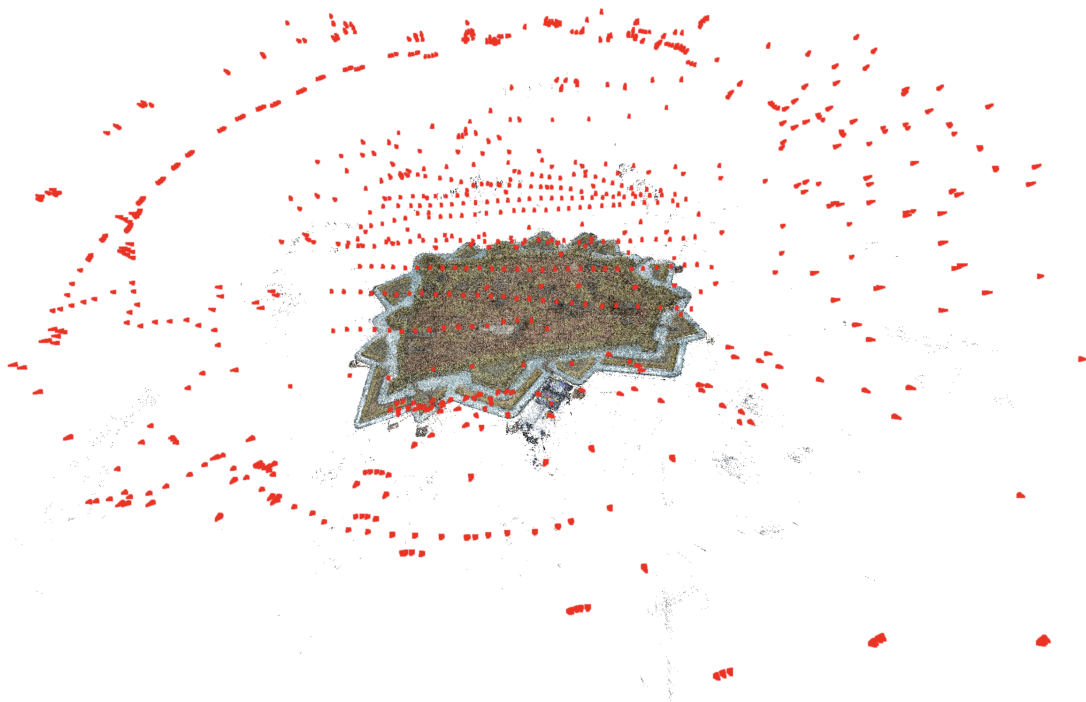


Figure 10. *Lys model* dataset : current "best-score" of TACO fully automated process on single iteration.

The aim of the proposed methodology is to contribute to photogrammetric-based documentation and analysis of CH objet providing a user-friendly tool (easily accessible through all-in-one command).

Operating from open-source solutions and compatible with remote-computing scenarios, the first results seem promising because only two automated steps allow to calculate a first 3D reconstruction on large and or complex data sets intended to be enriched by incremental registration of other imaging resources.

4. Conclusions and Perspectives

The processing methodology implemented and deployable through TACO offers a significative gain of robustness and versatility from FLOSS photogrammetric automated routines. The *Initial* module has proved interesting results to handle variated data sets in terms of size and complexity. The complementarity of BBA+SR approaches of the *Incremental* nodes provide an efficient co-registration framework consequently offering a great support to multi-modal CH imaging applications (including simultaneous multi-scale, multi-sensor, multi-temporal or multi- spectral management).

Initially developed as generic solution dedicated to remote-computing purposes, TACO shows interesting flexibility and scalability toward mixed modalities. The main successful condition still remains the photogrammetric correspondences and consistency (i.e., spatial and resolution gaps, spectral or radiometric deviations, temporal or contextual variations) in between the initial set and further the incremental ones. The recent works on absolute orientation and the interoperable functions offer new perspectives. Different side-projects of TACO would focus on extended spatial 2D/3D registration methods by creating a possible link with other modeling techniques (range-based, primitive-based etc).

Nonetheless, new challenging issues arose from the current achievements and will require a special care to add new features and optimize existing ones. On one hand, the added value in term of versatility, robustness and interoperability appears promising to build and enrich a digital twin by gathering variated image sources. On the other hand, the transition to a consistent object/space coordinate system will enable metric evaluation with other data flows. However reconstructing a 3D model from multiple sources necessitates a qualitative evaluation system to consolidate persistently their respective geometries (various density, quality, uncertainty, temporal stage). Indeed, it would be counterproductive to co-register indefinitely cameras and geometries without considering their positive or negative impact on merged outputs. Consequently, the registration of several geometries coming from different sources also increase the complexity of analysis while it shall instead be able to guide our interpretative process (i.e., by associating a data source or a modality to the observations it allows). Exploiting TACO to this end could provide an effective support toward data fusion (with a noticeable added value linking geometric features to their semantic attributes).

A new layer might have to be integrated in or on top of the proposed methodology to keep a trace of the processing steps while extracting indexes and metric useful to evaluate informative and geometric gains (or loss) of incremental and multimodal data fusion. Our latest experiments explored filtering approaches to reduce the redundancy of aggregated data to optimize and increase at the same time the interpretative value of long term incremental scenarios. In the meantime, advancements in CH-oriented approaches in un/supervised machine learning might provide interesting lead for classification [43] and segmentation purposes upstream [44] or downstream [45] TACO.

As a conclusion, if we anticipate the ongoing transition from accumulated one-shot applications toward a systemic approach of CH imaging, one might prevent the complexification of digital content management. In this scenario, the main challenge for the construction and the synchronism of CH-oriented digital twins will be our capacity to maintain methods and tools, enabling to manage and convey over-time, the growing richness of CH documentation.

Author Contributions: Conceptualization, A.P. and F.M.; methodology, A.P. ; software, F.M.; formal analysis, A.P. and F.M. ; writing—original draft preparation, A.P.; writing—review and editing, A.P., F.M., L.d.L. and P.V.; visualization, A.P. ; supervision, L.d.L. and P.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded supported by the ANR SUMUM project, grant Référence ANR-17-CE38-0004 of the French Agence Nationale de la Recherche.

Acknowledgments: The authors would like to acknowledge the stakeholders of the case-studies illustrating this paper, respectively the *Vasarely* and *Arbre-serpents* data sets correspond to the *Zett* painting of Victor Vasarely from the *Fondation Vasarely d'Aix-en-Provence* and the sculpture of Niki de Saint Phalle from the *Musée des Beaux-arts d'Angers*. The authors would also like to acknowledge their colleagues of FBK-3DOM in Trento and MAP-ARIA from Lyon for sharing respectively their data sets *Nettuno* and *Lys model* used and illustrated in this publication. A special acknowledgment to Éloi Gattet, president of the Mercurio start-up behind the idea and the draft of *Culbutto* module handling the QR code-based automatic scaling.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. AIOLI

AIOLI is a reality-based 3D annotation platform designed for a multidisciplinary CH community to construct semantically enriched descriptions of heritage assets from photogrammetric reconstruction and spatial labelling. This platform aims to impulse an innovative framework for the massive and large-scale collaborative CH documentation by linking features like Image-based modeling, 2D-3D spreading and correlation of semantic annotations, multi-layered analysis of qualitative and quantitative attributes. The service is deployed through a cloud-like infrastructure to make processing, visualization and annotation processes simply accessible via multi-device responsive web interfaces (PCs, tablets and smartphones) on-line and on-site. For more information on updates or to join beta-testing program, visit the website (<http://www.aioli.cloud/>).

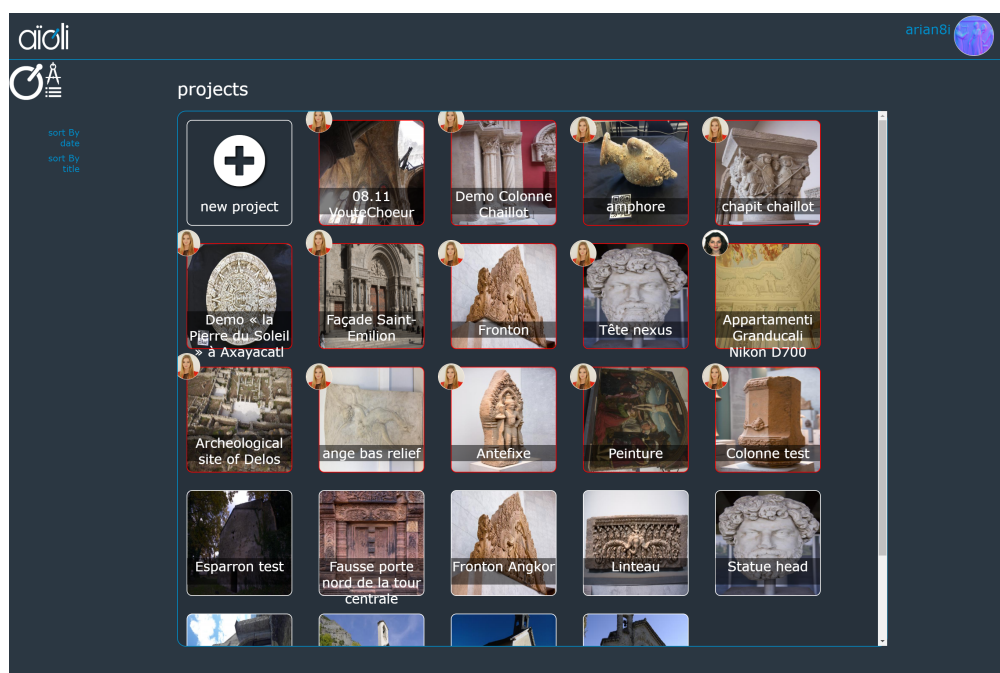


Figure A1. Current screenshot of AIOLI user's workspace showing the diversity of data sets processed by TACO.



Figure A2. *Saint-Trophisme* dataset : spatial-Resection of a web-based archives.

Cooperative Development between TACO and AIOLI

Once logged into AIOLI platform, users access a workspace to create a project (with public, shared or private visibility) linked to an Heritage Asset. The user is invited to upload an image set to obtain an initial photogrammetric reconstruction, powered by the fully automated TACO's workflow illustrated in Figure 2. The *Initial* process detailed in Section 2.2 aims to solve the self-optimization of the processing strategy from unknown inputs so as to increase the robustness and the relative accuracy expected to enable and support the semantic enrichment. This core feature of AIOLI is performed the semi-automatic 2D/3D annotation framework and the geometric analysis tools described in [46,47]. However, image-based documentation is no longer a linear workflow, nor one shot application, neither a punctual work done by a single end-user, with the same equipment, experience and purposes. The TACO *Incremental* processing modules detailed in Section 2.3 aim to support the collaborative documentation of CH objects implied by the growing complexity of multimodal scenarios currently performed in CH domain. Therefore the possibility to add and merge new image-sets through several iterations is simultaneously implemented within AIOLI and TACO in a cooperative framework. However initial cameras and geometry are not refined else all the annotations and their attributes must be recomputed. A new feature like the automated scaling and orientation based on coded-target detections was also initiated from the need to compute consistent coordinate system of AIOLI scenes (i.e images, annotations and descriptors). An other feature specially developed within TACO for AIOLI purposes is internet archive registration shown in Figure 6. Finally, the ongoing works on interoperability aim to create AIOLI's import/export functions for expanding its accessibility and the potentialities of collaborative annotations from and to other solutions.

References

1. Pamart, A.; Guillon, O.; Vallet, J.M.; De Luca, L. Toward a multimodal photogrammetric acquisition and processing methodology for monitoring conservation and restoration studies. In Proceedings of the 14th Eurographics Workshop on Graphics and Cultural Heritage, Genova, Italy, 5–7 October 2016; pp. 207–210.
2. Pamart, A.; Guillon, O.; Faraci, S.; Gattet, E.; Genevois, M.; Vallet, J.M.; De Luca, L. *Multispectral Photogrammetric Data Acquisition and Processing for Wall Paintings Studies*; ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences: Nafplio, Greece, 2017; pp. 559–566. [CrossRef]
3. Pamart, A.; Ponchio, F.; Abergel, V.; Alaoui M'Darhri, A.; Corsini, M.; Dellepiane, M.; Morlet, F.; Scopigno, R.; De Luca, L. A complete framework operating Spatially-Oriented RTI in a 3D/2D Cultural Heritage documentation and analysis tool. *Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci.* **2019**, XLII-2/W9, in press. [CrossRef]
4. Grussenmeyer, P.; Drap, P.; Gaillard, G. *ARPENTEUR 3.0: Recent Developments in Web based Photogrammetry*; ISPRS, Commission VI: Education and Communications: Sao Paulo, Brazil, September 2002; pp. 1–7.
5. Nilsson, D.; Pletinckx, D.; Van Gool, L.; Vergauwen, M. The ARC 3D Webservice. In *EPOCH Knowhow Book*; 2007; pp. 978–991; Available online: <http://www.her-it-age.net/> (accessed on 23 June 2020).
6. Heller, J.; Havlena, M.; Jancosek, M.; Torii, A.; Pajdla, T. 3D reconstruction from photographs by CMP SfM web service. In Proceedings of the 2015 14th IAPR International Conference on Machine Vision Applications (MVA), Tokyo, Japan, 18–22 May 2015; pp. 30–34.
7. Alliez, P.; Forge, F.; de Luca, L.; Pierrot-Deseilligny, M.; Preda, M. Culture 3D Cloud: A Cloud Computing Platform for 3D Scanning, Documentation, Preservation and Dissemination of Cultural Heritage. *ERCIM News* **2017**, 111, 64.
8. Nocerino, E.; Poiesi, F.; Locher, A.; Tefera, Y.T.; Remondino, F.; Chippendale, P.; Van Gool, L. 3D reconstruction with a collaborative approach based on smartphones and a cloud-based server. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2017**, XLII-2/W8, 187–194. [CrossRef]
9. Moulon, P.; Monasse, P.; Perrot, R.; Marlet, R. OpenMVG: Open multiple view geometry. In *International Workshop on Reproducible Research in Pattern Recognition*; Springer: Cham, Switzerland, 2016; pp. 60–74.
10. Rupnik, E.; Daakir, M.; Pierrot Deseilligny, M. MicMac—A free, open-source solution for photogrammetry. *Open Geospat. Data Softw. Stand.* **2017**, 2, 14 [CrossRef]
11. Hartley, R.I.; Mundy, J.L. Relationship between photogrammetry and computer vision. Integrating photogrammetric techniques with scene analysis and machine vision. *Int. Soc. Opt. Photonics* **1993**, 1944, 92–106.
12. Granshaw, S.I.; Fraser, C.S. Computer vision and photogrammetry: Interaction or introspection? *Photogramm. Rec.* **2015**, 30, 3–7. [CrossRef]
13. Aicardi, I.; Chiabrando, F.; Lingua, A.M.; Noardo, F. Recent trends in cultural heritage 3D survey: The photogrammetric computer vision approach. *J. Cult. Herit.* **2018**, 32, 257–266. [CrossRef]
14. Snavely, N.; Seitz, S.M.; Szeliski, R. Modeling the world from internet photo collections. *Int. J. Comput. Vis.* **2008**, 80, 189–210. [CrossRef]
15. Triggs, B.; McLauchlan, P.F.; Hartley, R.I.; Fitzgibbon, A.W. Bundle adjustment—A modern synthesis. In *International Workshop on Vision Algorithms*; Springer: Heidelberg, Berlin, 1999; pp. 298–372.
16. Remondino, F.; Fraser, C. Digital camera calibration methods: Considerations and comparisons. *Int. Arch. Photogramm. Remote. Sens.* **2006**, 36, 266–272.
17. Gonzalez-Aguilera, D.; López-Fernández, L.; Rodriguez-Gonzalvez, P.; Hernandez-Lopez, D.; Guerrero, D.; Remondino, F.; Menna, F.; Nocerino, E.; Toschi, I.; Ballabeni, A.; et al. GRAPHOS—open-source software for photogrammetric applications. *Photogramm. Rec.* **2018**, 33, 11–29. [CrossRef]
18. Gonzalez-Aguilera, D.; Ruiz de Ona, E.; López-Fernandez, L.; Farella, E.M.; Stathopoulou, E.K.; Toschi, I.; Remondino, F.; Rodríguez-Gonzálvez, P.; Hernández-López, D.; Fusiello, A.; et al. PhotoMatch—An open-source multi-view and multi-modal feature matching tool for photogrammetric applications. *ISPRS Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2020**, in press.
19. Remondino, F.; El-Hakim, S. Image-based 3D Modelling: A Review. *Photogramm. Rec.* **2006**, 21, 269–291, doi:10.1111/j.1477-9730.2006.00383.x. [CrossRef]

20. Remondino, F.; Nocerino, E.; Toschi, I.; Menna, F. A critical review of automated photogrammetric processing of large datasets. In Proceedings of the 26th International CIPA Symposium 2017, Ottawa, ON, Canada, 28 August–1 September 2017; Volume 42, pp. 591–599.
21. Stathopoulou, E.K.; Welponer, M.; Remondino, F. Open-Source Image-Based 3d Reconstruction Pipelines: Review, Comparison and Evaluation. *Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci.* **2019**, *42*, 331–338. [[CrossRef](#)]
22. Barazzetti, L.; Remondino, F.; Scaioni, M.; Lo Brutto, M.; Rizzi, A.; Brumana, R. Geometric and radiometric analysis of paintings. *Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci.* **2010**, *38*, 62–67.
23. Nocerino, E.; Rieke-Zapp, D.H.; Trinkl, E.; Rosenbauer, R.; Farella, E.; Morabito, D.; Remondino, F. Mapping VIS and UVL imagery on 3D geometry for non-invasive, non-contact analysis of a vase. *Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci.* **2018**, *42*, 773–780. [[CrossRef](#)]
24. Grifoni, E.; Legnaioli, S.; Nieri, P.; Campanella, B.; Lorenzetti, G.; Pagnotta, S.; Poggialini, F.; Palleschi, V. Construction and comparison of 3D multi-source multi-band models for cultural heritage applications. *J. Cult. Herit.* **2018**, *34*, 261–267. [[CrossRef](#)]
25. Chane, C.S.; Mansouri, A.; Marzani, F.S.; Boochs, F. Integration of 3D and multispectral data for cultural heritage applications: Survey and perspectives. *Image Vis. Comput.* **2013**, *31*, 91–102. [[CrossRef](#)]
26. Yeh, C.K.; Matsuda, N.; Huang, X.; Li, F.; Walton, M.; Cossairt, O. A streamlined photometric stereo framework for cultural heritage. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 738–752.
27. Van der Perre, A.; Hameeuw, H.; Boschloos, V.; Delvaux, L.; Proesmans, M.; Vandermeulen, B.; Van Gool, L.; Watteeuw, L. Towards a combined use of IR, UV and 3D-Imaging for the study of small inscribed and illuminated artefacts. In *Lights On... Cultural Heritage and Museums!*; LabCR Laboratory for Conservation and Restoration, FLUP Faculty of Arts and Humanities, University of Porto: Porto, Portugal, 2016; pp. 163–192.
28. Suwardhi, D.; Menna, F.; Remondino, F.; Hanke, K.; Akmalia, R. Digital 3D Borobudur-Integration of 3D surveying and modeling techniques. *Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci.* **2015**, *40*, 417. [[CrossRef](#)]
29. Ramos, M.M.; Remondino, F. Data fusion in cultural heritage a review. *ISPRS–Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci.* **2015**, *5*, W7. [[CrossRef](#)]
30. Toschi, I.; Capra, A.; De Luca, L.; Beraldin, J.A.; Cournoyer, L. On the evaluation of photogrammetric methods for dense 3D surface reconstruction in a metrological context. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, *II-5*, 371–378, doi:10.5194/isprsannals-ii-5-371-2014. [[CrossRef](#)]
31. Calvet, L.; Gurdjos, P.; Griwodz, C.; Gasparini, S. Detection and accurate localization of circular fiducials under highly challenging conditions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 562–570.
32. Wang, J.; Olson, E. AprilTag 2: Efficient and robust fiducial detection. In Proceedings of the IEEE/RISJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Korea, 9–14 October 2016.
33. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]
34. Rosu, A.M.; Assenbaum, M.; De la Torre, Y.; Pierrot-Deseilligny, M. Coastal digital surface model on low contrast images. *Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci.* **2015**, *40*, 307. [[CrossRef](#)]
35. Hartmann, W.; Havlena, M.; Schindler, K. Recent developments in large-scale tie-point matching. *ISPRS J. Photogramm. Remote. Sens.* **2016**, *115*, 47–62. [[CrossRef](#)]
36. Farella, E.; Torresani, A.; Remondino, F. Sparse point cloud filtering based on covariance features. *Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci.* **2019**, *42*, 465–472. [[CrossRef](#)]
37. Shih, T.; Faig, W. A solution for space resection in closed form. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **1988**, *27*, 547–556.
38. Larsson, V.; Kukulova, Z.; Zheng, Y. Making minimal solvers for absolute pose estimation compact and robust. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2335–2343.
39. Ke, T.; Roumeliotis, S.I. An efficient algebraic solution to the perspective-three-point problem. *arXiv* **2017**, arXiv:1701.08237.

40. Persson, M.; Nordberg, K. Lambda Twist: An Accurate Fast Robust Perspective Three Point (P3P) Solver. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 318–332.
41. Palma, G.; Corsini, M.; Dellepiane, M.; Scopigno, R. Improving 2D-3D Registration by Mutual Information using Gradient Maps. In Proceedings of the Eurographics Italian Chapter Conference, Genova, Italy, 18–19 November 2010; pp. 89–94. [[CrossRef](#)]
42. Merkel, D. Docker: Lightweight Linux Containers for Consistent Development and Deployment. *Linux J.* **2014**, *2014*, 2.
43. Aceto, G.; Ciunzo, D.; Montieri, A.; Pescapè, A. MIMETIC: Mobile encrypted traffic classification using multimodal deep learning. *Comput. Networks* **2019**, *165*, 106944. [[CrossRef](#)]
44. Anzid, H.; Le Goic, G.; Bekkari, A.; Mansouri, A.; Mammass, D. Multimodal Images Classification using Dense SURF, Spectral Information and Support Vector Machine. *Procedia Comput. Sci.* **2019**, *148*, 107–115. [[CrossRef](#)]
45. Belhi, A.; Bouras, A.; Foufou, S. Leveraging known data for missing label prediction in cultural heritage context. *Appl. Sci.* **2018**, *8*, 1768. [[CrossRef](#)]
46. Manuel, A.; Véron, P.; De Luca, L. 2D/3D semantic annotation of spatialized images for the documentation and analysis of cultural heritage. In Proceedings of the 14th Eurographics Workshop on Graphics and Cultural Heritage, Genova, Italy, 5–7 October 2016; pp. 101–104.
47. Manuel, A.; M'Darhri, A.A.; Abergel, V.; Rozar, F.; De Luca, L. A semi-automatic 2D/3D annotation framework for the geometric analysis of heritage artefacts. In Proceedings of the 2018 3rd Digital Heritage International Congress (DigitalHERITAGE), San Francisco, CA, USA, 26–30 October 2018; pp. 1–7.

Sample Availability: Data sets presented in this paper are available and shareable on request, in case there is no conflict with the right-holders of the artworks. TACO library will be released as a deliverable of the ANR SUMUM project, grant Référence ANR-17-CE38-0004 of the French Agence Nationale de la Recherche.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).