A Comparative Analysis of Different Software Packages for 3D Modelling of Complex Geometries

Styliani Verykokou[1(✉)], Sofia Soile[1], Fotis Bourexis[1], Panagiotis Tokmakidis[2], Konstantinos Tokmakidis[2], and Charalabos Ioannidis[1]

[1] School of Rural and Surveying Engineering, NTUA, 9 Iroon Polytechniou, 15780 Athens, Greece
st.verykokou@gmail.com, {ssoile,cioannid}@survey.ntua.gr, fotis.bourexis@gmail.com
[2] School of Rural and Surveying Engineering, AUTH, 54124 Thessaloniki, Greece
ptokmaki@gmail.com, ktok@auth.gr

Abstract. The purpose of this paper is the investigation of the performance of four well-established commercial and open-source software packages for automated image-based 3D reconstruction of complex cultural and natural heritage sites, i.e., Agisoft Metashape, RealityCapture, MicMac and Meshroom. The case study is part of the inaccessible giant rock of St. Modestos, in the archaeological site of Meteora. In terms of computational time, the commercial software packages were the most time-efficient solutions, with Metashape being the fastest one. They also have a friendlier user interface, which makes them adoptable even by non-photogrammetrists. All four solutions yielded approximately comparable results in terms of accuracy and may be used for generation of 3D dense point clouds of complex sites. With the exception of Meshroom, they may produce georeferenced results. Also, with the exception of MicMac, which did not yield satisfactory results in terms of textured mesh, they may be used for generating photorealistic 3D models. The comparative analysis of the results achieved by the tested software will serve as the basis for establishing photogrammetric pipelines that may be generally used for 3D reconstruction of complex geometries.

Keywords: 3D model · Geometric documentation · Software evaluation

1 Introduction

The importance of 3D documentation of cultural and natural heritage sites is well-understood at an international level and experts attempt to use modern technologies to produce highly accurate and detailed 3D models of such sites. Several works have been conducted in recent years, showing promising results achieved via photogrammetric methods, using images [1] or combination of images and laser scanning techniques [2]. Some cultural and natural sites correspond to complex geometries, either because they are inaccessible or because of their magnitude and geometric characteristics. Thus, their

M. Ioannides et al. (Eds.): EuroMed 2020, LNCS 12642, pp. 228–240, 2021.
https://doi.org/10.1007/978-3-030-73043-7_19

3D modelling requires specific attention. Such a site is the UNESCO world heritage site of Meteora, characterized by inaccessible giant rocks with morphological peculiarities and challenging topographical features. The 3D geometric documentation of Meteora, which is dealt with within the ongoing "METEORA" project [3], is a highly demanding task, accomplished using images from unmanned aerial vehicles (UAVs) and manned aircrafts, terrestrial images, LiDAR data and ground control points (GCPs) [4].

The purpose of this paper is the investigation of the performance of well-established commercial (Agisoft Metashape [5] and RealityCapture [6]) and open-source (MicMac [7] and Meshroom [8]) software packages for the automated 3D reconstruction of complex cultural and natural sites. The study area is part of the rock of St. Modestos, known as "Modi", located in the Meteora site. On top of this rock, ruins of the old monastery of St. Modestos exist. It is of great height (about 200 m) and the ascent to this rock is of increased difficulty, so it was covered by UAV images. Its topographic features are representative of complex cultural and natural sites, so it was selected as the study area.

2 Image-Based 3D Modelling

The reconstruction of the 3D scene geometry from images is a problem that has occupied the photogrammetric community for more than 40 years. Advances in photogrammetry and computer vision have led to the development of automated structure from motion (SfM) and multi-view stereo (MVS) approaches that have seen tremendous evolution over the years. SfM refers to the process of estimating the camera poses corresponding to a 2D image sequence and reconstructing the sparse scene geometry [9]. MVS is the general term given to a group of methods using stereo correspondences as their main cue in more than two images [10]. The combination of SfM and MVS provides automated workflows for generating dense 3D point clouds and surface models.

The first step of SfM is the extraction of features in each image [11]; SIFT-based algorithms are the most commonly used ones. The matching of the descriptors is the next step, using the criterion of a minimum distance measure, followed by outlier rejection techniques. The correspondences are then organized into tracks [9]. An incremental, hierarchical or global method follows. Incremental methods register one camera at each iteration; hierarchical ones gradually merge partial reconstructions; and global methods register all cameras simultaneously [12]. In case of incremental and hierarchical methods, intermediate bundle adjustment processes are necessary to ensure successful camera pose estimation and sparse 3D point cloud extraction, in addition to a final bundle adjustment, as required by global methods. Georeferencing of the SfM results is generally performed via a 3D similarity transformation between the arbitrary SfM system and the world reference system using GCPs and/or GPS measurements.

The generation of a dense point cloud is the next step within a MVS scheme, through a dense image matching (DIM) algorithm, using either a stereo (via a local, global or semi-global algorithm) or a multi-view approach. Local methods compute the disparity at a given point using the intensity values within a finite region, thus trading accuracy for speed. Global ones are more accurate but time consuming; they solve a global optimization by minimizing a cost function based on the whole image. Semi-global methods perform a pixel-wise matching, allowing to shape efficiently object boundaries and details, and represent a good trade-off between accuracy and speed [13]. The conversion of dense cloud into mesh and its texturing are the final steps of the MVS pipeline.

3 Experiments

3.1 Test Dataset

A dataset consisting of 238 UAV images depicting part of the giant rock of St. Modestos, known as Modi, was used in the experiments. The rock is located in the archaeological site of Meteora, in central Greece, near the town of Kalambaka. Meteora hosts one of the largest and most precipitously built complexes of Eastern Orthodox monasteries. Modi features a complex geometry, being a giant inaccessible rock with challenging topography. The images were captured by a DJI camera using a DJI Phantom 4 Pro UAV. They correspond to a size of 5,472 × 3,648 pixels, a focal length of 8.8 mm and a pixel size of 2.41 μm. They are accompanied by GPS/INS information.

The ground coordinates of 6 GCPs were computed via Agisoft Metashape using a georeferenced model of the Meteora site. The latter was generated using aerial images of Meteora and GCPs in the wider area. The geometry of Modi did not permit the on-site measurement of GCPs, so the computed coordinates of these 6 GCPs were used for georeferencing. All experiments were performed using a 64-bit Intel Core i7-8700 CPU 3.2 GHz computer with 24 GB of RAM and MS Windows 10 Pro operating system.

3.2 Agisoft Metashape

Agisoft Metashape [5], developed by Agisoft LLC., is a commercial software that generates 3D models from images. Its pipeline consists of four fully automated steps, i.e., SfM, DIM, meshing and texturing, which let the user set various parameters, along with some optional steps, e.g., manual measurement of GCPs or tie points, manual or semi-automatic definition of masks, etc. It has a very simple graphical user interface and offers a high degree of automation, making it usable even by less experienced users.

The first was the alignment of the images. It is a SfM process that uses the available GPS/INS data to generate a georeferenced sparse point cloud of the scene and compute the camera poses and optionally their interior orientation. A modification of SIFT is used for feature extraction. The feature points were extracted in images of original size. Thresholds of 30,000 features and 15,000 tie points per image were specified, so that the sparse point cloud does not consist of too many points and the alignment process is not computationally intensive. The image pairs used for matching were selected using the GPS/INS data, to avoid matching of all possible pairs. Camera calibration was performed during the alignment step, using a frame camera model. A distortion model encompassing 11 degrees of freedom (DoF) was used: 1 for focal length, 2 for principal point, 2 for affinity and skew transformation coefficients, 4 for radial distortion coefficients and 2 for tangential distortion coefficients. The alignment time was 16 min.

6 GCPs were measured in the corresponding images, resulting in 129 image measurements, and their ground coordinates were inserted (modified, up to a translation transformation, to overcome a visualization issue in the derived point clouds in case of big coordinates). The points were added manually as “markers”. The optimization of the cameras took place, via auto-calibrating bundle adjustment for exterior orientation estimation and sparse cloud generation in the reference system defined by the GCPs.

DIM was the next step. Metashape calculates depth information for each camera and combines it into a single dense point cloud. DIM was performed using the "medium" quality setting, which implies image downscaling by a factor of 16 (4 times by each side). The "aggressive" depth filtering mode was used to sort out most of the outliers. The computational time of the dense point cloud generation process was 2 h 36 min.

The dense cloud was then transformed into mesh. The "arbitrary" surface type was chosen and the maximum number of polygons was set to 1/5 of the number of dense cloud points, via the "high" setting. Image downscaling by 4 (2 times by each side) was selected via the "high" quality setting. In order to automatically fill holes in areas without points, interpolation mode was enabled, according to which Metashape interpolates some surface areas within a circle around every point. The meshing lasted 17 min.

The generation of texture was the last step. The "generic" mapping mode, which does not make any assumptions on the scene type, and the "mosaic" mode, which performs blending of the low frequency component for overlapping images and uses the high frequency one from one image, were used. Hole filling and ghosting filter were disabled. The texture size was set to 15,000 × 15,000 pixels. This step lasted 10 min.

3.3 RealityCapture

RealityCapture [6] is a commercial software package developed by Capturing Reality s.r.o., which generates 3D models from images, laser scans or combination of both. Its pipeline consists of alignment, reconstruction and texturing. By the term reconstruction, RealityCapture implies both the DIM and meshing processes.

The alignment was the first step. Although its documentation is limited, probably a modified SIFT algorithm is used. The alignment mode was set to "high", i.e., a setting targeted to highly overlapping images; RealityCapture was set to detect 60,000 features per image and keep 10,000 of those for further matching and processing. This initial alignment was completed in 4 min, using the GPS/INS metadata for georeferencing.

The same GCPs measured within Metashape were added in the corresponding images within RealityCapture. Then, using the update alignment tool, information about the reprojection error of each GCP was available. Running the final alignment was much faster, because there was no need for running SIFT again; RealityCapture keeps this information from the initial alignment. The estimated values of exterior and interior orientation were more precise, compared to the corresponding values of the initial alignment. The distortion model used was "Brown3 with Tangential2" that has 8 DoF: 1 for focal length, 2 for principal point, 3 for radial distortion coefficients and 2 for tangential distortion coefficients. The final alignment was completed in 2 min.

Reconstruction was the next step for creating a 3D mesh of the surveyed area. "Normal" mode was selected, without any downscaling of the images, and specifying a maximum of 5,000,000 vertices per part and a detail decimation factor of 1, indicating no decimation for smoothing details when creating the mesh. The unwrapping style was set to "maximal texture count"; unwrapping parameters were set to: 8,192 × 8,192 pixels; optimal texel size was calculated to 0.0083 m and set to 0.016 m for processing. A decimation took place within RealityCapture using the "simplify" command to reduce the number of triangles from 153.2M to 50M (21.5M vertices). The processing times were as follows: depth mapping: 1 h 9 min; meshing: 2 h 46 min; post-processing: 9 min.

RealityCapture provides a set of tools for selecting vertices and filtering them out, i.e., the reconstruction region bounding box, a lasso, rectangle and box tool for 3D selection, and the "Advanced" selection tool, which calculates the average edge length and provides four different selection options: marginal triangles; largest connected component; small triangles; and large triangles using a given threshold by (times × average edge length). Those tools were used to clean up the model and the topology of the model was checked (check topology tool) before proceeding to texturing.

The final step was texturing. The model was textured using multiple texture files (18 texture images) of 8,192 × 8,192 pixels resolution and the processing time was 21 min.

3.4 MicMac

MicMac [7] is a free open-source photogrammetric suite developed by IGN, France, which can be used for image-based 3D reconstruction. It consists of a set of command line tools, permitting a high degree of parameterization. Some visual interfaces are also provided, by calling the appropriate command, to facilitate parameter tuning. The target users are rather professionals, with a basic knowledge of photogrammetry.

The SfM procedure was completed using eight MicMac commands and seven of its tools. Initially, the OriConvert tool was used for transformation of the GPS/INS data accompanying the images from text format to MicMac's orientation format and generation of a file with the pairs of overlapping images. This task was completed in 1.5 min.

The Tapioca tool was used for computation of tie points using SIFT in images of original size. The file exported by OriConvert was used as input in Tapioca, so that tie points are extracted only in overlapping images. A first experiment was applied in images of original size, resulting in 13 h 30 min of computational time. Due to the extremely long processing time of Tapioca for full-resolution images, a second experiment was conducted for extracting feature points in images downscaled by 16, i.e., 4 times by each side (1,368 × 912 pixels). The processing time was dramatically different, as it took only 27.5 min. The fact that Tapioca does not provide the possibility of adjusting SIFT thresholds to extract a maximum number of features per image, e.g., like Metashape, makes it computationally ineffective in case of full-resolution images. For instance, about 750,000 features were extracted per image in the first experiment, while the maximum number of features per image in Metashape was set equal to 40,000.

The Tapas tool was used for camera calibration and relative orientation. "RadialStd" mode was selected, indicating an 8DoF distortion model: 1 for focal length, 2 for principal point, 2 for distortion center and 3 for coefficients of radial distortion. The processing time was 23 h 57 min using the tie points of the first test, whereas the Tapas command was completed in 14 h 28 min using the tie points of the second test.

A sparse cloud of the scene including the camera poses, in an arbitrary coordinate system, was created via AperiCloud in 23.5 min for the first test and 3.5 min for the second one. Whereas the output of this step is not used in any subsequent tool, it is useful for visualization reasons. Until this step, the GPS/INS values by MicMac are only used for image pairs determination, without being used for georeferencing reasons.

Whereas MicMac has a tool for measuring GCPs (SaisieAppuisInitQT), it was not used, in order to apply the same measurements made via Metashape. The coordinates of the GCPs, exported by Metashape, were converted in formats readable by MicMac.

The transformation of the relative orientation, as computed by Tapas, into absolute orientation was performed via the GCPBascule tool, using as input the image and ground coordinates of the GCPs. This command was completed in 5 s for both tests.

The bundle adjustment of the whole block of images was conducted via Campari. Self-calibration was not performed within the block adjustment. Whereas Campari provides the option of using GPS values within the adjustment, they were not used in the experiments. This step lasted 2 h 43 min for the first test and 11 min for the second one.

The creation of a sparse cloud of the scene including the camera poses in the ground reference system was implemented via AperiCloud. The processing times were similar to the ones achieved before absolute orientation and bundle adjustment, i.e., 25 min for the first test and 2.5 min for the second one. This was the last step of the SfM procedure.

The dense point cloud was created through automated DIM via the C3DC tool. The "BigMac" option was used, according to which the 3D coordinates of 1 point per 4 pixels are computed through DIM. A color point cloud was the output of this process. DIM lasted 18 h 21 min for the first test and 2 h 15 min for the second one.

In order to visualize the point cloud and crop it, so that it depicts the geometry of the area of interest, the MeshLab software [14] was used, as MicMac does not provide any tool for visualization and editing of 3D models. MeshLab is a free open-source 3D mesh processing software, developed by the ISTI-CNR institute of Italy. The cropping process was manual and was applied for the dense clouds generated by both tests.

The generation of a 3D mesh was implemented via the MicMac tool TiPunch, using the cropped dense cloud of each test via the Poisson reconstruction algorithm. The maximum reconstruction depth was set to 8, as any higher setting regarding a bigger reconstruction depth was too computationally ineffective. Meshing took 4 min for each test.

The Tequila tool was used for texturing, using the "Stretch" criterion for selecting the best image for each triangle, i.e., the best stretching of triangle projection in the image. The "Angle" criterion that takes into account the angle between the triangle normal and image viewing direction was also tested but was discarded, as it produced worse results. "Basic" mode was used, according to which all images are stored in the texture map. The maximum texture dimension was set to 15,000 pixels. It lasted 5 min.

3.5 Meshroom

Meshroom [8] is a free open-source 3D reconstruction software based on the AliceVision framework that produces textured models and provides its users with the possibility of parameterizing each of its steps. Once the parameterization is specified, the whole processing may be automatically completed. Meshroom permits input of additional images, while the processing is ongoing. Also, it can perform a live reconstruction. However, it does not provide the possibility of adding GCPs. It requires CUDA-enabled GPU, with a computing capability of at least 2.0. Its photogrammetric pipeline includes two main stages, i.e., SfM and MVS, and eleven basic steps, referred to as nodes.

Within the SfM stage, the camera intrinsic parameters were loaded from the image metadata and SIFT feature extraction took place. "Low quality" was selected, taking into account the quality and viewing angles of the cameras. This process took 3 min. A quick (<10 s) image matching preprocessing step was applied for determining the image pairs, without the cost of resolving all matches in detail, through tree classification. Then, the main process of image matching took place, followed by RANSAC for outlier rejection. This process took 3 min. An incremental SfM method was used for computing the camera poses and generating a sparse point cloud, which was completed in 2 min.

The undistortion of the images through the PrepareDenseScene node was then implemented in less than 5 min. The DepthMap and DepthMapFilter nodes were two of the most time consuming procedures (38 h without image downscaling); they were applied in order to retrieve the depth value of each pixel for all cameras and force depth consistency, respectively. As soon as these steps were finished, the dense point cloud and the arbitrary polygon mesh were generated through the Meshing node in 2 h. Values such as the maximum number of points of the point cloud, observation angle and factor, etc. were specified. The noise of the primary polygon mesh was largely eliminated by the MeshFiltering node, in which a smoothing operation took place (1.5 min), preparing the polygon mesh for texturing. Attribute values, such as the unwrapping mode, the resolution, etc. were user-specified. A maximum texture size of 8,192 pixels was specified for texturing, which took no more than 20 min without image downscaling. The dense point cloud and mesh model were cropped via the Meshlab software.

Table 1 outlines the tools of the tested software concerning each stage of 3D modelling and the basic parameterization selected. Table 2 indicates the computational time for each test.

Table 1. Tools of the tested software and parameterization used in each one

Stage	Metashape	RealityCapture	MicMac	Meshroom
Search for pairs	**Align Photos** (GPS/INS use; full-resolution images; feature point limit: 30,000; tie point limit: 15,000; 11DoF distortion model)	**Align** (GPS/INS use; alignment mode: high; max features per image: 60,000; preselector features: 10,000; 8DoF distortion model)	**OriConvert** (GPS/INS use)	**Image Matching**
Feature extraction and matching			**Tapioca** (test 1: full-resolution images; test 2: images downscaled by 16	**Feature Extraction** (downscaled by 16); **Feature Matching**
Interior and relative orientation			**Tapas** (8DoF distort. Model)	**Structure-from-Motion**
Sparse point cloud			**AperiCloud**	
Measurement of GCPs	**Markers** (6 GCPs measured: 129 image measurements in total)	**Markers** (the GCPs measured in Metashape were used)	**SaisieAppuisInitQT** (the GCPs measured in Metashape were used)	-
Absolute orientation	**Optimize Cameras** (autocalibration)	**Align** (alignment before and after GCPs input)	**GCPBascule**	
Bundle adjustment			**Campari** (no autocalibration)	
Sparse point cloud			**AperiCloud**	

(*continued*)

Table 1. (*continued*)

Stage	Metashape	RealityCapture	MicMac	Meshroom
DIM	**Build Dense Cloud** (downscaled by 16; aggressive depth filtering)	**Reconstruction** (detail level: normal; no image downscaling; max vertices count per part: 5,000,000; no decimation; no editing of dense point cloud)	**C3DC** (the 3D coordinates of 1 point per 4 pixels are computed)	**PrepareDenseScene; DepthMap; DepthMapFilter; Meshing**
Cropping of dense point cloud	**Free-form selection**		**MeshLab** (no editing tools)	**MeshLab** (no editing tools)
Generation of 3D mesh	**Build Mesh** (arbitrary surface; high face count; interpolation enabled)		**TiPunch** (Poisson reconstruction; max reconstruction depth = 8)	**Meshing**
Cropping of 3D mesh	**Free-form selection**	**Selection toolbox; Simplification tool**	**MeshLab** (no editing tools)	**MeshLab** (no editing tools)
Generation of textured 3D model	**Build Texture** (mode: generic, mosaic; size: 15,000)	**Texture** (visibility-based; size: 8,192; max count: 40)	**Tequila** (criter.: Stretch; mode: Basic; size: 15,000)	**Texturing** (mode: basic; max size: 8,192)

Table 2. Computational time of the tests implemented using the four software packages

	Metashape	RealityCapture	MicMac test 1	MicMac test 2	Meshroom
SfM	0 h 16 min	0 h 6 min	41 h 0 min	15 h 14 min	0 h 8 min
DIM	2 h 36 min	1 h 8 min	18 h 21 min	2 h 15 min	38 h 0 min
Meshing-texturing	0 h 27 min	3 h 24 min	0 h 9 min	0 h 9 min	2 h 18 min
Total time	3 h 19 min	4 h 38 min	59 h 30 min	17 h 38 min	40 h 26 min

4 Results

The main results are summarized in Table 3. Meshlab does not provide any information concerning the number of tie points (matches). The number of tie points per image was set to be fixed in the case of RealityCapture. In the other two software packages, the average, maximum and minimum number of matches were different. The average number of tie points was too big in the case of the first MicMac test using the full-resolution images for feature extraction, as it does not provide the possibility of defining an upper threshold. Thus, the only available solution to reduce the computational time was to downscale the images, as implemented in the second test. Metashape displays the number of matches per image but may not export them. Hence, whereas such statistics could also be estimated for Metashape as well, they require a great deal of manual processing, so their extraction was discarded. The maximum number of tie points per image was quite similar for the case of Metashape, RealityCapture and the second test of MicMac.

Table 3. Main results of the experiments conducted using the tested software solutions

Metric	Metashape	RealityCapture	MicMac test 1	MicMac test 2	Meshroom
Avg tie points per image	n/a	10,000	181,295	11,030	n/a
Max tie points per image	11,200	10,000	283,023	13,991	n/a
Min tie points per image	n/a	10,000	49,845	3,584	n/a
Avg residual of tie points	0.58 pix	0.38 pix	0.40 pix	0.83 pix	1.10 pix
Max residual of tie points	43.45 pix	0.99 pix	0.46 pix	1.03 pix	4.00 pix
Avg GCPs residual	0.57 m	n/a	0.52 m	0.53 m	n/a
Max GCPs residual	0.69 m	n/a	0.74 m	0.71 m	n/a
Avg GCPs residual in axial components X, Y, Z (m)	0.21, 0.25, 0.43	0.22, 0.25, 0.38	0.24, 0.25, 0.34	0.23, 0.24, 0.37	n/a
Max GCPs residual in axial components X, Y, Z (m)	0.40, 0.36, 0.60	0.42, 0.45, 0.55	0.46, 0.56, 0.56	0.42, 0.53, 0.60	n/a
RMS error of GCPs in axial components X, Y, Z (m)	0.23, 0.30, 0.45	n/a	0.27, 0.30, 0.38	0.25, 0.29, 0.41	n/a
Sparse cloud points	0.65M	4.7M	35.6M	2.7M	0.018M
Dense cloud points	24.2M	81.9M	24.4M	24.8M	24.0M
Dense cloud points (cropped)	21.2M	25.1M	21.9M	21.9M	12.8M
Vertices of final 3D mesh	1.4M	25.1M	0.07M	0.07M	8.2M
Faces of final 3D mesh	2.9M	50.0M	0.15M	0.11M	13.4M

The smallest average residual of tie points was observed for RealityCapture and the biggest one for Meshroom. The first MicMac test yielded the smallest worst residual. The worst residual of tie points for Metashape was quite big, indicating that at least an

outlier was not removed. However, this does not generally influence the rest Metashape results, which are satisfying.

Regarding the average residual of GCPs in axial components [X, Y, Z], approximately equivalent results were reported by Metashape, RealityCapture and MicMac. The X and Y residuals and RMS errors were similar for Metashape and MicMac, whereas the Z residual and RMS error were the worst for Metashape. Similar results were reported for the RMS errors of GCPs in axial components for Metashape and MicMac, while this kind of information was not available for RealityCapture and Meshroom.

The sparse cloud density was significantly lower for Meshroom and Metashape. This number for the first MicMac test was too big, due to lack of any tie points threshold. The dense cloud density of Metashape, MicMac and Meshroom was comparable. Whereas the number of dense cloud points after editing within RealityCapture was equivalent with the other solutions, its initial dense cloud included 3 times more points. The meshing and texturing MicMac tools are still under development, so the textured MicMac meshes were not satisfactory, as shown in Fig. 1. The rest textured models are visually satisfying. The numbers of mesh vertices and faces are significantly lower for MicMac. The biggest numbers of vertices and faces were reported by RealityCapture.

The computational time of these software packages differs a lot. Metashape is the quickest option, as the total processing time was less than 3.5h, while RealityCapture is also very fast, completing the 3D reconstruction process in a little more than 4.5h.

Furthermore, comparisons in the derived dense point clouds were made using the free open-source CloudCompare software [15], via its "Cloud to Cloud Distance" tool. The Metashape dense cloud was assumed to be the reference one. While the RealityCapture and MicMac point clouds were georeferenced, the Meshroom cloud was in an arbitrary system; hence, it was aligned to the reference one via measurement of common points, followed by the ICP algorithm. The mean and standard deviation of distances are presented in Table 4. The smallest mean difference was observed for the first test of MicMac (full-resolution images for alignment) and the biggest one for its second test. Comparable results are derived using all software packages, as verified by the mean differences between these dense clouds, which do not exceed 7.5 cm. The order of magnitude of these differences is quite smaller than the uncertainty of the models in the reference system, which is quite big, due to the quality of GCPs (see Sect. 3.1) An interesting aspect is the fact that the MicMac dense clouds derived using different alignment parameterization yield comparable differences from the reference Metashape cloud; hence, a computationally intensive full-resolution matching via MicMac is not generally needed, taking into account the time parameter. Figure 2 provides a visualization of the absolute differences (m) between the Metashape dense cloud and each one of the four compared clouds. The largest differences are observed in the edges of all dense clouds, due to the insufficient number of overlapping images depicting these regions.

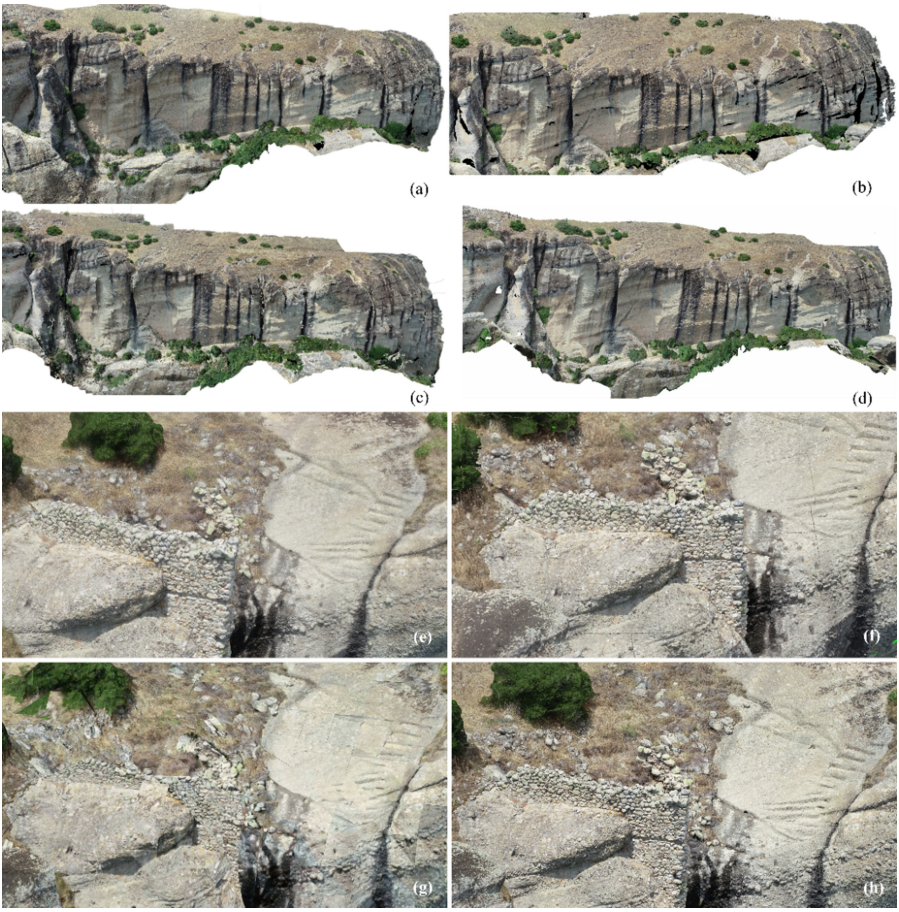

Fig. 1. Textured 3D models and zoom-in views derived using Metashape (a, e), RealityCapture (b, f), MicMac - test 2 (c, g) and Meshroom (d, h)

Table 4. Distances between the reference (Metashape) dense cloud and the compared ones

Metric	RealityCapture	MicMac - test 1	MicMac - test 2	Meshroom
Mean (cm)	6.9	5.8	7.4	6.7
Std. Dev. (cm)	9.4	6.4	7.5	8.5

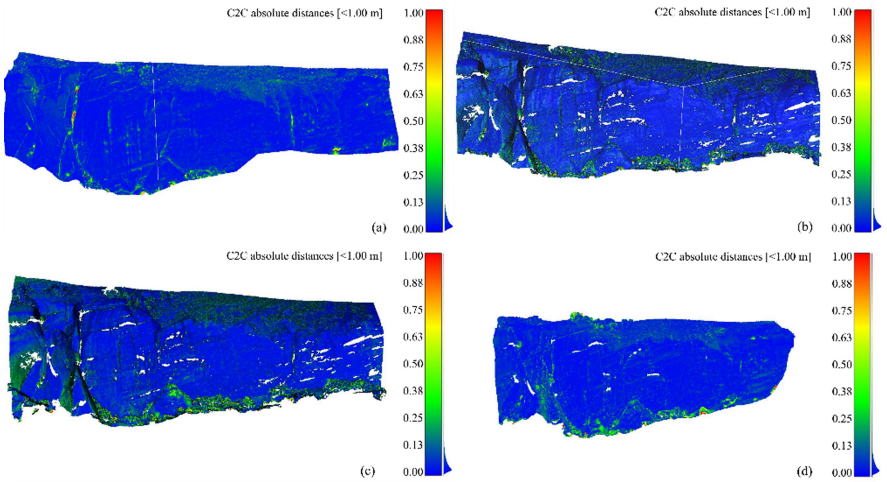

Fig. 2. Absolute differences (m) between the reference (Metashape) and the compared dense point clouds, i.e., RealityCapture (a), MicMac - test 1 (b), MicMac - test 2 (c) and Meshroom (d).

5 Discussion and Conclusions

Within this paper, the performance of four well-established commercial and free software solutions was evaluated for image-based reconstruction of complex cultural and natural heritage sites. Regarding bundle adjustment results, all solutions produced comparable outputs in terms of accuracy (taking into account tie points and GCP residuals as well as RMS errors, where applicable). The mean distance of the derived dense point clouds is almost negligible, whereas biggest differences are observed in the edges of the dense clouds. A major disadvantage of Meshroom was the fact that it does not provide the possibility for measuring GCPs; hence, its results refer to an arbitrary coordinate system. MicMac produced satisfactory results in terms of dense point cloud; however, its final textured mesh model was not satisfactory, as the corresponding tools are still under development. The investigation of the use of the MicMac dense cloud for mesh generation via another software solution, e.g., MeshLab, and its texturing either using MicMac or another software using the orientation of images produced by MicMac would be interesting. In terms of computational time, the commercial software packages were the most efficient solutions, with Metashape being the fastest one. The commercial software have a friendlier user interface, which makes them adoptable even by non-photogrammetrists. Also, Meshroom is quite user-friendly, giving the possibility of quite wide parameterization. On the other hand, MicMac consists of command line tools which can be used by experts in photogrammetry, thus not being easy to use.

In conclusion, in cases of geometric documentation of complex sites in a ground system defined by GCPs, Metashape and RealityCapture are suitable for generating a textured 3D surface model, while MicMac is suitable for generating a 3D dense point cloud, which may be inserted in another software for the meshing and texturing process. Meshroom may only be used for generating a 3D model in an arbitrary coordinate system. On the other hand, Metashape and RealityCapture are commercial software, so

if the budget of an organization or project does not permit a purchase of their licenses, both free solutions yield acceptable results in terms of accuracy and dense point clouds. Their combination with a mesh processing software would probably produce satisfactory results; this is an issue that will be investigated within our future research.

Acknowledgements. This research has been co-financed by the European Union and Greek national funds through the Operational Program Competiveness, Entrepreneurship and Innovation, under the call RESEARCH–CREATE–INNOVATE (project code: T1EDK02859).

References

1. Verykokou, S., Doulamis, A., Athanasiou, G., Ioannidis, C., Amditis, A.: Multi-scale 3D modelling of damaged cultural sites: use cases and image-based workflows. In: Ioannides, M., et al. (eds.) Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection, vol. 10058, pp. 50–62. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-48496-9_5
2. Jo, Y., Hong, S.: Three-dimensional digital documentation of cultural heritage site based on the convergence of terrestrial laser scanning and unmanned aerial vehicle photogrammetry. ISPRS Int. J. Geo-Inf. **8**(2), 53 (2019). https://doi.org/10.3390/ijgi8020053
3. METEORA project. https://www.meteora.net.gr/. Accessed 01 Sept 2020
4. Ioannidis, C., Verykokou, S., Soile, S., Boutsi, A.-M.: A multi-purpose cultural heritage data platform for 4D visualization and interactive information services. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. **XLIII-B4-2020**, 583–590 (2020). https://doi.org/10.5194/isprs-archives-XLIII-B4-2020-583-2020
5. Agisoft Metashape. https://www.agisoft.com/. Accessed 01 Sept 2020
6. Capturing Reality. https://www.capturingreality.com/. Accessed 01 Sept 2020
7. MicMac Wiki. https://micmac.ensg.eu/index.php/Accueil. Accessed 01 Sept 2020
8. AliceVision. https://alicevision.org/. Accessed 01 Sept 2020
9. Verykokou, S., Ioannidis, C.: A photogrammetry-based structure from motion algorithm using robust iterative bundle adjustment techniques. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. **IV-4/W6**, 73–80 (2018). https://doi.org/10.5194/isprs-annals-IV-4-W6-73-2018
10. Furukawa, Y., Hernández, C.: Multi-view stereo: a tutorial. Found. Trends Comput. Graph. Vis. **9**(1–2), 1–148 (2015). https://doi.org/10.1561/0600000052
11. Tareen, S.A.K., Saleem, Z.: A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK. In: iCoMET 2018, pp. 1–10. IEEE (2018)
12. Verykokou, S., Ioannidis, C.: Exterior orientation estimation of oblique aerial images using SfM-based robust bundle adjustment. Int. J. Remote Sens. **41**, 7217–7254 (2020)
13. Nalpantidis, L., Christou Sirakoulis, G., Gasteratos, A.: Review of stereo vision algorithms: from software to hardware. Int. J. Optomechatron. **2**(4), 435–462 (2008)
14. MeshLab. https://www.meshlab.net/. Accessed 01 Sept 2020
15. CloudCompare. https://www.danielgm.net/cc/. Accessed 01 Sept 2020