3D Documentation and Risk Assessment of Sicilian Fortified Heritage: Integrated Low-Cost UAV Imagery and Videogrammetric Protocols

Antonella Versaci[1], Alessio Cardaci[2,3]
[1] Dept. of Engineering and Architecture, University of Enna 'Kore', Enna, Italy - antonella.versaci@unikore.it
[2] Dept. of Engineering and Applied Sciences, University of Bergamo, Bergamo, Italy - alessio.cardaci@unibg.it
[3] Institute for Construction Technologies (ITC), Italian National Research Council (CNR), L'Aquila, Italy - alessio.cardaci@unibg.it

Keywords: SfM-MVS; 3D image-based modelling; non-professional videos; user-generated visual sources; masonry decay; conservation planning

Abstract

Sicilian fortified heritage is a rich and widespread system, yet at the same time fragile and exposed to risk conditions determined by the geomorphological complexity of the sites, local slope instability, erosion, difficult accessibility and the progressive decay of masonry structures. Castles, towers and defensive outposts are often located on elevated ground, urban margins, steep slopes, rocky outcrops or coastal fronts, in close relation to the historic landscape and to territorial settings that are now subject to increasing vulnerability. These conditions require rapid, sustainable and replicable documentation procedures, capable of supporting knowledge of the monument and a preliminary assessment of conservation issues. The paper proposes an integrated low-cost protocol for preliminary 3D documentation and for supporting the risk assessment of this valuable heritage. The protocol combines low-cost UAV imagery, non-professional video sequences, critical frame selection, SfM-MVS photogrammetry, data quality control and the production of knowledge-oriented three-dimensional models. Particular attention is given to the reuse of images and videos produced by non-specialist users, a vast, still underexploited visual archive that, if interpreted through explicit procedures of selection and validation, can support the 3D documentation of monuments at different stages of transformation. The experimentation concerns three case studies: Castelluccio di Gela, Castello di Agira and Rocca di Aci Castello. These were selected because they represent three different conditions, allowing the transferability of the workflow and the usefulness of 3D models to be tested for the preliminary reading of structures, decay phenomena and vulnerability conditions.

1. Introduction

The fortified heritage of Sicilian fortified heritage forms a fragile architectural and territorial system, exposed to environmental pressures arising from the interaction of geomorphological complexity, slope instability, erosive processes, masonry decay and the progressive loss of structural legibility. Castles, towers and defensive outposts are often located on rocky outcrops or steep slopes, where the relationship with orography was an essential component of military function and territorial control. The same condition that underpins their historical and landscape value has today become a source of vulnerability, as limited accessibility, vegetation, fragmentary remains and the difficulty of working safely complicate survey, monitoring and preliminary damage assessment. These architectures derive their importance not from the quality of isolated monuments but from their belonging to a complex settlement and geographical system, constructed over time through visual, defensive, administrative and landscape relationships (Bresc and Maurici, 2009). The rapid and integrated cataloguing of Sicilian fortified heritage therefore becomes a central issue and must be addressed in relation to hazard quantification, highlighting the need for knowledge tools capable of connecting measurements, accessibility, vulnerability and planned conservation (Versaci et al., 2020). Up-to-date documentation of the condition of the structures forms the basis for scientific investigation capable of integrating architectural form, material conditions, morphological context and conservation issues. Reading the relationship between masonry structure, ground, urban margins and the natural environment is therefore essential for guiding subsequent investigations, defining intervention priorities and developing tools for the management of precarious conditions. Image-based 3D reconstructions now play a significant role in the protection and valorisation of built heritage (Federman et al., 2017). In fortified sites, aerial acquisition makes it possible to record wall tops, curtain walls, towers, changes in level and surfaces that are difficult to observe from the ground, improving the continuity of representation and the understanding of the relationship between architecture and site (Fallavollita and Ugolini, 2017). This approach is part of a broader research framework devoted to technological innovation for the study, restoration and promotion of the island's cultural heritage, in which digital survey is understood not as a simple tool for representation, but as a critical device for interpreting the artefact, its context and its conservation conditions (Cardaci and Versaci, 2019). Accurate metric surveys, especially when carried out using aerial systems, require appropriate equipment, specific expertise and careful planning of the data-acquisition phases. This makes systematic and timely work on a rich and extensive cultural resource difficult, especially within the framework of a planned maintenance programme. Today, however, alongside acquisitions produced through professional campaigns, a widespread, fragmentary and still underused visual repertoire has developed. The growing use of commercial drones, smartphones, lightweight video cameras and sharing platforms has made available an increasing number of non-professional images and videos of the anthropised landscape: tourist photographs, occasional footage, unplanned UAV sequences and online content. Although not recorded for geometric purposes, these sources may contain valuable information on vulnerable, poorly accessible or insufficiently mapped structures. In some cases, they constitute the only visual evidence of conditions that can no longer be observed; in others, they document minor heritage assets overlooked by systematic measurement campaigns. Their value lies in the possibility of transforming multimedia content into a tool for understanding the artefact. Studies conducted on visual sources from social media have shown how images and sequences not originally produced for survey purposes can be reused to build 3D models of cultural sites (Themistocleous, 2017). Similarly, research on photographic collections available online has demonstrated the possibility of processing large sets of photographs produced by different users in order to highlight, within three-dimensional scenes, the transformations undergone by monuments over time (Martin-Brualla et al., 2015). This suggests that non-professional archives may constitute a significant resource for field investigation, provided that they are subjected to procedures of selection, control and critical interpretation. This does not mean that amateur videos, rapid sequences and low-cost UAV images are equivalent to a planned geometric framework. The methodological issue concerns the criteria within which such sources can contribute to the generation of digital replicas useful for critical analysis, morphological reading and continuous updating of the heritage asset. For the assessment of vulnerabilities in defensive complexes, these 3D models make it possible to observe masonry continuity, gaps, collapses, invasive vegetation, geometric discontinuities, relationships with the ground and degrees of accessibility. The result is an initial information base useful for recognising instability, planning targeted inspections and guiding subsequent surveys, without replacing specific structural, geotechnical or diagnostic analyses. This contribution proposes an integrated, low-cost protocol for the 3D archiving of Sicilian military architecture and for supporting risk assessment. The aim is to define the operational conditions, potential and limits of the informed reuse of a heritage corpus of considerable historical and cultural relevance and varied complexity. The experimentation concerned three case studies: Castelluccio di Gela, Castello di Agira and Rocca di Aci Castello. The three case studies make it possible to assess the transferability of the protocol in different territorial scenarios and the usefulness of low-cost 3D models as preliminary tools for knowledge, reading the condition of sites and planning subsequent investigations (fig. 1).

2. Videogrammetry as a strategy for 3D reconstruction

Image-based, reality-based surveying now plays a central role in processes of knowledge, conservation and management of fortified heritage, particularly in the case of rugged, fragmentary and poorly accessible sites. Structure from Motion (SfM) and Multi-View Stereo (MVS) procedures, now consolidated in photogrammetric practice, make it possible to estimate camera orientation and scene geometry from partially overlapping images, generating, in sequence, sparse clouds, dense models, textured meshes and orthoprojections (Schönberger and Frahm, 2016). Videogrammetry sits within this framework as an operational method capable of transforming a video sequence into a photogrammetric dataset. A series of sequential frames can be extracted from a video and, if critically selected, treated as partially overlapping images. The temporal continuity of video often favours a high degree of overlap between contiguous frames, facilitating the recognition of homologous points and the alignment of the sequence. A comparison between photogrammetry from single UAV images and videogrammetry from video frames highlights specific advantages and limitations. Photogrammetry from individual images allows greater control over image acquisition, which can be defined according to the geometry of the artefact, flight altitude, acquisition angle, overlap and lighting conditions. The quality of the individual image is generally higher than that of video frames, because the image is less compressed and exposure, focus and sensitivity parameters are more controllable. Videogrammetry, by contrast, makes it possible to obtain a larger number of frames that, although of lower quality, can partly compensate for the weaknesses of individual frames through greater observational redundancy. This makes it particularly suitable, where extreme accuracy is not required, for the rapid documentation of paths, curtain walls, rock faces, irregular surfaces and parts of the artefact that are difficult to record through isolated photographic stations. It also allows the reuse of materials already available, such as amateur clips, previous footage or content published online, thereby extending the documentation of sites lacking planned metric campaigns (Alsadik, 2016). The limitations of videogrammetry are nevertheless significant and must be considered during the selection and validation process. Frames may be affected by low resolution, high compression, motion blur, rolling shutter, automatic exposure and unstable focus. These factors degrade the sharpness of individual frames, reduce the recognisability of features and may affect the final accuracy of the reconstruction (Sun and Zhang, 2019). The trajectories of non-professional videos are often discontinuous: the camera moves through approaches, retreats, sudden rotations, interruptions and uncontrolled variations in framing. When movement occurs predominantly along the optical axis of the camera, with a

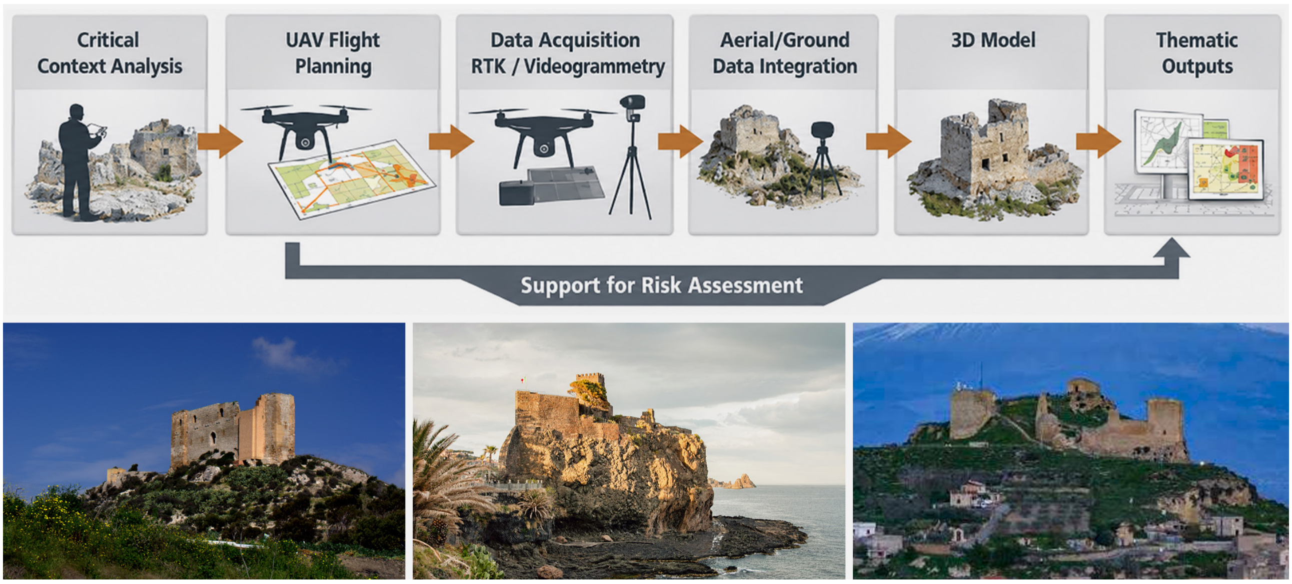

Figure 1. Phases of the 3D documentation workflow for Sicilian fortified heritage: Castelluccio di Gela, Rocca di Aci Castello and Castello di Agira.

limited lateral component, parallax is reduced, triangulation becomes weaker and uncertainty in depth estimation increases. Benchmarks conducted on historical footage confirm that the metric accuracy of the reconstruction varies considerably according to the type of camera motion used, generally being more favourable in movements that introduce a significant lateral component than in purely axial ones. Frame selection is therefore a crucial methodological step: the automatic extraction of an excessive number of images, or even of all available images, may increase computational load and introduce redundancies that are not always useful, without necessarily improving reconstruction stability. The construction of the dataset must instead derive from a case-by-case assessment, according to the visual quality of the frames, the continuity of camera motion, the overlap between consecutive images and the presence of geometrically significant information. The operator interprets the available material and defines a selection strategy aimed at privileging sharp frames that are sufficiently differentiated and capable of describing the object surfaces progressively. The reliability and cost-effectiveness of videogrammetry are closely dependent on the definition of explicit criteria for selection, control and validation. Some studies have explored automatic selection procedures based on the regularisation of the camera path and on the evaluation of the geometric quality of individual frames, with the aim of reducing the load of the subsequent bundle adjustment phase without compromising the redundancy required for reconstruction (Pavoni et al., 2016). Other contributions have tested the use of automatic recognition and neural networks to isolate, within historical or heterogeneous footage, the portions of architectural interest to be used in the photogrammetric process, thereby reducing manual intervention during dataset preparation (Condorelli and Rinaudo, 2019). The literature therefore points to an evolving field, in which automatic and semi-automatic tools may influence the phases of selection, organisation and pre-processing of visual data. The present contribution is situated within this methodological context but emphasises the critical control of the operator as a central factor. Amateur video sequences are not treated as immediately usable materials, but as documentary sources to be subjected to reasoned selection, controlled processing and coherence checks.

3. Dataset processing and optimisation

The experimentation began with the review of amateur videos available online, not originally produced for measurement purposes but potentially reusable for 3D documentation. This phase involved critical analysis and selection, aimed at transforming heterogeneous, discontinuous and uncontrolled visual materials into image sequences compatible with photogrammetric processing (fig. 2). The selection of footage was not based solely on general visual quality, but prioritised verification of camera movements, trajectory continuity, the persistence of the artefact within the frame and the possibility of ensuring sufficient overlap between consecutive frames (Westoby et al., 2012). Radiometric quality and image legibility are, in fact, essential characteristics for reconstructing point positions and for the geometric coherence of the model (Cardaci et al., 2024).

A crucial step concerned the choice of image-extraction frequency, conducted not automatically but in relation to the specific characteristics of each sequence: movement speed, distance from the artefact, camera stability, presence of rotations, scale variations, focus quality and legibility of the architectural object. Not all available frames were included, since an excessive density of similar images generates redundancy, increases processing times and does not necessarily improve the metric or descriptive quality of the model. The sampling frequency was therefore defined case by case, with the aim of obtaining an overlap of between 70% and 80%. In this phase, the operator's role is decisive, because the construction of the dataset requires a technical assessment rooted in the specific expertise of the architectural surveyor. It is necessary to recognise which portions of the footage can contribute to block stability and which, by contrast, risk introducing noise, ambiguity or geometric weakness into the network. The choice of frame step was therefore based on the assessment of visual continuity, image overlap, recognisability of homologous points and overall stability of the acquisition network (tab. 1). Attention to the spatial structure of the acquisitions is particularly important in the case of amateur videos, where the camera trajectory is often random and irregular. It is not sufficient to have many frames; each must contribute to the stability of the system. In the presence of short-baseline images, increasing the number of points used for orientation has a limited impact on precision if the geometry is weak (Barazzetti, 2017); likewise, configurations with almost parallel and only slightly divergent acquisition axes may favour systematic errors (James and Robson, 2014). The operator must therefore assess not only the sharpness of the single shot but also its geometric usefulness, excluding portions of video characterised by stationary flight, abrupt rotations, sudden scale variations, significant occlusions, blurred acquisitions or insufficient visibility of the artefact. Cuts and extractions were oriented towards constructing robust datasets, even when obtained from uncontrolled sources. The procedure therefore took on a critical character, selecting within the stream only those sequences compatible with the logic of this methodological approach. The video clips relating to the three case studies were chosen because they had the same resolution (1920 × 1080 px) and the same

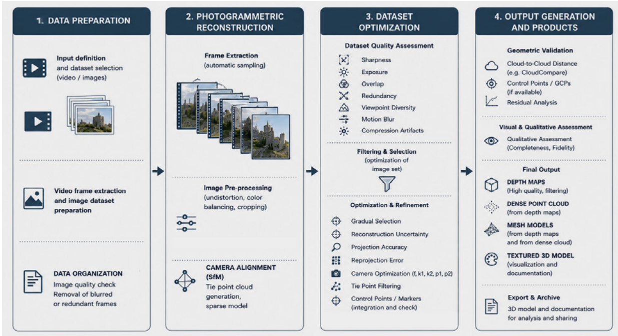

Figure 2. Videogrammetric workflow: video selection, frame extraction, SfM-MVS processing, optimisation and 3D outputs.

Case studies / video characteristics	Castelluccio of Gela	Rocca of Aci Castello	Castle of Agira
Total video duration	1 min 14 s	1 min 53 s	8 min 48 s
Resolution	1920 x 1080 px	1920 x 1080 px	1920 x 1080 px
Image format	16:9 Full HD	16:9 Full HD	16:9 Full HD
Original video FPS	30	30	30
Video codec	H.264 / AVC	H.264 / AVC	AVC / H.264 L4.1
Video bitrate	approx. 3.88 Mb/s	approx. 4.38 Mb/s	14.7 Mb/s
Extracted segment duration	63.90 s	45.15 s	449 s / 7 min 29 s
Frame step (1 frame every ...)	5	3	16
Number extracted frames	384	452	858
Extraction frequency (fps)	approx. 6 fps	approx. 10 fps	approx. 1.90 fps
Average interval between frames	0.167 s	0.100 s	0.52 s
Estimated apparent GSD	5-15 cm/px	4-12 cm/px	3-8 cm/px
Main dataset characteristic	Intermediate density; good readability in views	Highest-density dataset; subject clearly in the frame	Longest dataset, but less dense and more heterogeneous

Table 1. Technical characteristics of the videos used for the case studies.

Parametri	Castelluccio of Gela				Rocca of Aci Castello				Castle of Agira			
	Raw Sparse Cloud	Optimized Sparse	Absolute Variation	Percentage Variation	Raw Sparse Cloud2	Optimized Sparse	Absolute Variation2	Percentage Variation2	Raw Sparse Cloud3	Optimized Sparse	Absolute Variation5	Percentage Variation6
Images	384	384	**0**	**0,0%**	446	446	**0**	**0,0%**	850	850	**0**	**0,0%**
Tie points total	448.129	448.129	**0**	**0,0%**	459.808	459.808	**0**	**0,0%**	1.129.075	1.129.075	**0**	**0,0%**
Tie points used	252.304	160.966	**-91.338**	**-36,2%**	236.832	113.001	**-123.831**	**-52,3%**	804.552	485.459	**-319.093**	**-39,7%**
RMS reprojection error (pix)	0,920076	0,486724	**-0,433352**	**-47,1%**	0,485639	0,328948	**-0,156691**	**-32,3%**	1,128830	0,770731	**-0,358099**	**-31,7%**
Max reprojection error (pix)	40,100100	5,280200	**-34,819900**	**-86,8%**	19,564000	3,248020	**-16,315980**	**-83,4%**	41,067100	6,175460	**-34,891640**	**-85,0%**
Mean key point size (pix)	3,338020	2,402020	**-0,936000**	**-28,0%**	3,446620	2,615310	**-0,831310**	**-24,1%**	3,608850	2,401140	**-1,207710**	**-33,5%**
Average tie point multiplicity	5,50496	5,50496	**0,00000**	**0,0%**	8,09585	8,09585	**0,00000**	**0,0%**	7,00570	7,00570	**0,00000**	**0,0%**
Matching time	2 min 52 s	2 min 52 s	**0 min 00 s**	**0,0%**	3 min 04 s	3 min 04 s	**0 min 00 s**	**0,0%**	8 min 13 s	8 min 13 s	**0 min 00 s**	**0,0%**
Alignment time	6 min 06 s	6 min 06 s	**0 min 00 s**	**0,0%**	12 min 45 s	12 min 45 s	**0 min 00 s**	**0,0%**	40 min 33 s	40 min 33 s	**0 min 00 s**	**0,0%**

Table 2. Alignment and Tie Point Cloud optimisation parameters before and after filtering for the three case studies.

frame rate (30 fps), although they differed considerably in duration and bit rate. This led to different choices in extraction frequency: from the video of the Castelluccio di Gela, one frame was extracted every 0.16 sec; from that of the Rocca di Aci Castello, one frame every 0.10 sec; and from that of Castello di Agira, the most extensive and characterised by the highest bit rate, one frame every 0.52 sec. The sequences also allowed an indicative estimate of the Ground Sample Distance (GSD), derived from the GNSS data and EXIF metadata of the UAV systems. The resulting value cannot be considered highly accurate, since reliable data on acquisition distances, effective focal length, sensor size and optical-axis orientation were not available. Nevertheless, the GSD proved useful as a preliminary comparative indicator for assessing the documentary potential of the different datasets, while remaining a purely indicative value and not a metric parameter of absolute accuracy. The three datasets were processed in the same way so that they could be compared despite differences in the number of images, the geometric complexity of the sites and the acquisition conditions. The methodological choice was to keep the main alignment parameters unchanged in order to assess the response to the same processing strategy; for all datasets, alignment was performed at High Accuracy, maintaining high values for Key Point Limit and Tie Point Limit, respectively 1,000,000 and 10,000,000, with a ratio of 1:10. This setting was adopted to extract a large number of correspondences from the photographs, without allowing the software to select only the points it considered best. The validity of this choice was confirmed by the complete alignment of all images in the three datasets, which demonstrated that the blocks possessed sufficient internal coherence to position all cameras correctly in space (tab. 2). The Tie Point Cloud thus obtained, however, contains both geometrically robust points and less reliable points, generated by weak correspondences, blurred images, reduced parallax and local matching errors. For this reason, progressive and selective filtering is required through Gradual Selection, alternating the controlled removal of less stable points with camera optimisation. Camera optimisation consists of recalculating the block through bundle adjustment, with refinement of the internal and external orientation parameters and of the triangulated coordinates of the Tie Points. The first filter concerned Reconstruction Uncertainty, which is useful for identifying points reconstructed with weak triangulation geometry, due to a reduced baseline or to an unfavourable image distribution. The second filter concerned Projection Accuracy, which is useful for identifying points whose projection is poorly localised because of keypoint detection within the image-resolution pyramid, since points extracted from coarser levels are less reliable. The filtering and optimisation sequence was repeated three times, with controlled point reductions of approximately 10%-15% at each step, so as to progressively improve the coherence of the Tie Point Cloud without excessively impoverishing the observation network. The filter based on Reprojection Error, which measures the distance between the observed position of a point in the image and the position recalculated by the photogrammetric model, was applied only in the final phase, since excessively early or aggressive filtering

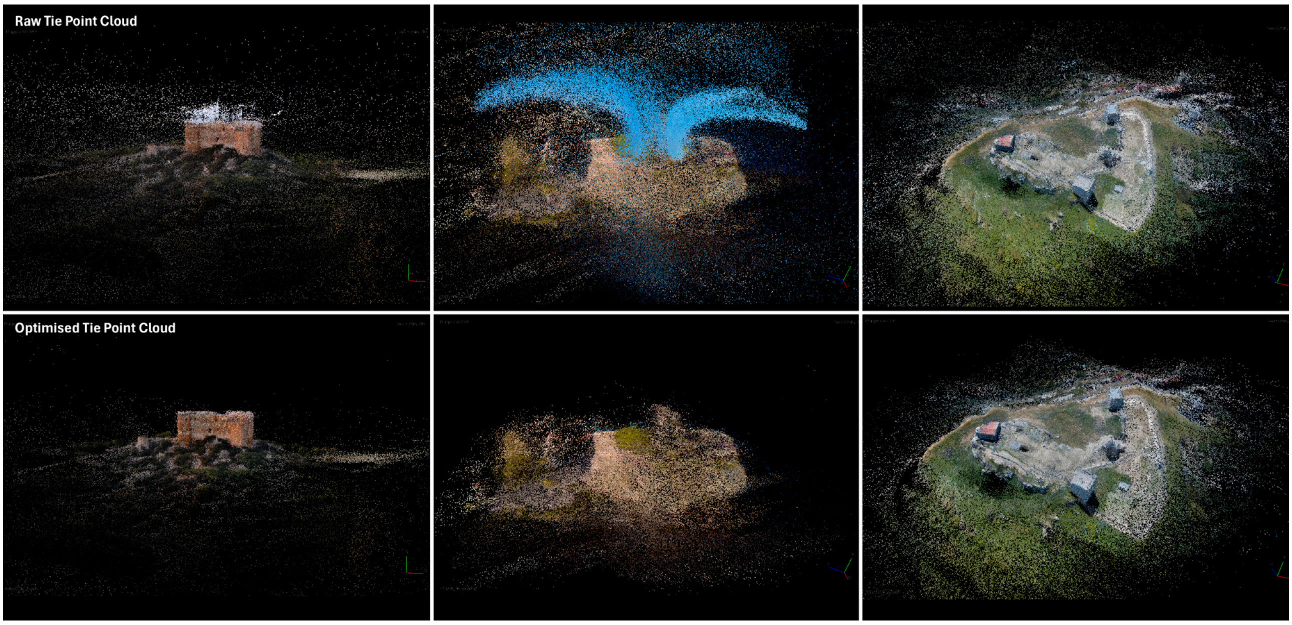

Figure 3. Tie Point Clouds (before and after filtering and optimisation) for Castelluccio di Gela, Rocca di Aci Castello and Castello di Agira.

could have reduced the robustness of the network. The effectiveness of the optimisation was verified through quantitative and visual-qualitative checks aimed at assessing both the extent of the correction and the density and spatial distribution of the points. The quantitative assessment considered the number of Tie Points Used, the RMS Reprojection Error, the Max Reprojection Error, the Mean Key Point Size and the Average Tie Point Multiplicity. These parameters describe, respectively, the number of points actually used in orientation, the mean reprojection residual, the maximum residual value, the average size of the detected features and the redundancy of observations. Since Ground Control Points or homogeneous direct measurements were not available for the three case studies, the evaluation was conducted as an internal coherence check of the process. It was based on the alignment parameters, the spatial distribution of the Tie Points, the completeness of the reconstructed surfaces and the comparison between derived outputs. The process operated systematically on the three blocks, with different intensities according to the initial conditions of the datasets. All the main indicators showed an improvement: the Tie Points Used were substantially reduced, the RMS Reprojection Error decreased despite starting from initially limited values, and the Max Reprojection Error showed the most marked contraction, with reductions of more than 83% in the three case studies. This trend indicates that the procedure acted above all on the removal of the most unstable points, which were responsible for the highest residuals. The Mean Key Point Size also decreased, confirming the selection of features that were, on average, smaller and more precisely localisable than those in the initial set. The Average Tie Point Multiplicity, by contrast, remained substantially stable; observational redundancy was therefore not compromised by optimisation, and overlap between images retained a structuring role in the robustness of the network. The results demonstrated that optimisation improved the reliability, geometric distribution and overall coherence of the blocks and, above all, made it possible to process different datasets according to a comparable and replicable logic (fig. 3). The procedure adopted therefore makes it possible to transform heterogeneous datasets into more controlled photogrammetric blocks, laying the foundations for the subsequent production of three-dimensional models (Dense Cloud and Mesh) adequate for preliminary documentation and for reading the vulnerabilities of Sicilian fortified heritage. The traceability of operational choices and the verification of processing parameters are essential and unavoidable conditions for using 3D data as a critical support for conservation (Cardaci et al., 2026). This aspect is particularly relevant in the case of low-cost video sequences and images, where the variability of the acquisitions may generate redundant, noisy and unstable observations.

4. Strategies for generating 3D models

Once the Tie Point Cloud had been optimised, 3D model generation was carried out in three consecutive phases: the generation of Depth Maps, the production of first-level outputs derived directly from them, and the derivation of further models from these outputs (fig. 4). The abbreviations used below distinguish both the level of derivation and the nature of the product: the letter A indicates first-level outputs, obtained directly from the Depth Maps; the letter B indicates second-level outputs, derived from the A outputs. The number 1 identifies meshes, while the number 2 identifies point clouds.

Phase 1 (generation of Depth Maps). The first step following optimisation concerned the generation of Depth Maps, calculated from the aligned images and the camera orientation parameters. Depth Maps are 2.5D images that associate each pixel with a distance from the camera, adding depth information to the two-dimensional representation. This processing was carried out by setting the Quality parameter to High for all three datasets, in order to maintain a high level of detail in depth estimation, and by using the Mild filter to obtain non-aggressive noise reduction compatible with the preservation of detail.

Phase 2 (first-level outputs from Depth Maps). Starting from the Depth Maps, two main first-level outputs were generated: the Mesh from Depth Maps (A1) and the Dense Cloud from Depth Maps (A2). The two products derive from the same information base but reconstruct the geometry of the scene according to different logics. The Mesh from Depth Maps (A1) was obtained by using the depth maps directly as the source, without passing through the preliminary generation of the Dense Cloud; the process builds a continuous polygonal surface from the depth information calculated on the images. The result is a more compact, homogeneous and manageable model, in which some of the redundancies and micro-incoherences of the dense data are reduced. Its main advantage is the lightness of the model, in terms of both face count and overall file size, and therefore its greater manageability; the limitation is that the surface may be more regularised, attenuating small edges, minor discontinuities or local irregularities. The Dense Cloud from Depth Maps (A2), by contrast, is the photogrammetric dense cloud generated by the fusion of the Depth Maps; unlike the Mesh, it describes the surface through a large set of points with coordinates and colour. This output preserves a larger amount of point-based information and is therefore more suitable for subsequent editing, filtering, classification, export and analysis operations. The advantage of the Dense Cloud from Depth Maps (A2) lies in the greater descriptive richness of the data, which makes it possible to work on point distribution and to intervene in a more controlled way on any noisy, lacunar or redundant areas. The disadvantage concerns file size, inhomogeneity and the possible presence of noise.

Phase 3 (second-level outputs). The third phase involved the production of two further outputs derived from the previous ones: the Mesh from Dense Cloud (B1) and the Point Cloud from Mesh (B2). These products are not a simple duplication of the previous processes but represent a further mode of data refinement: the Mesh from Dense Cloud (B1) is generated from the Dense Cloud from Depth Maps (A2); the Point Cloud from Mesh (B2), by contrast, is obtained by sampling the Mesh from Depth Maps (A1). It is important to distinguish the specific behaviour of the two second-level outputs in relation to the first-level ones. The Mesh from Dense Cloud (B1) has a more irregular surface and more vertices because it directly inherits the distribution, density

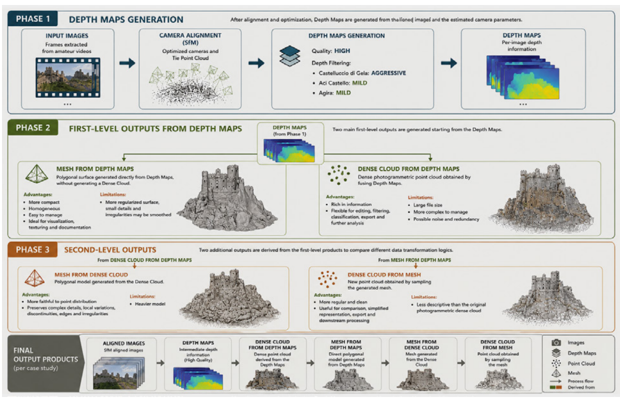

Figure 4. Videogrammetric workflow: first-level from Depth Maps, second-level from model generated in the previous step.

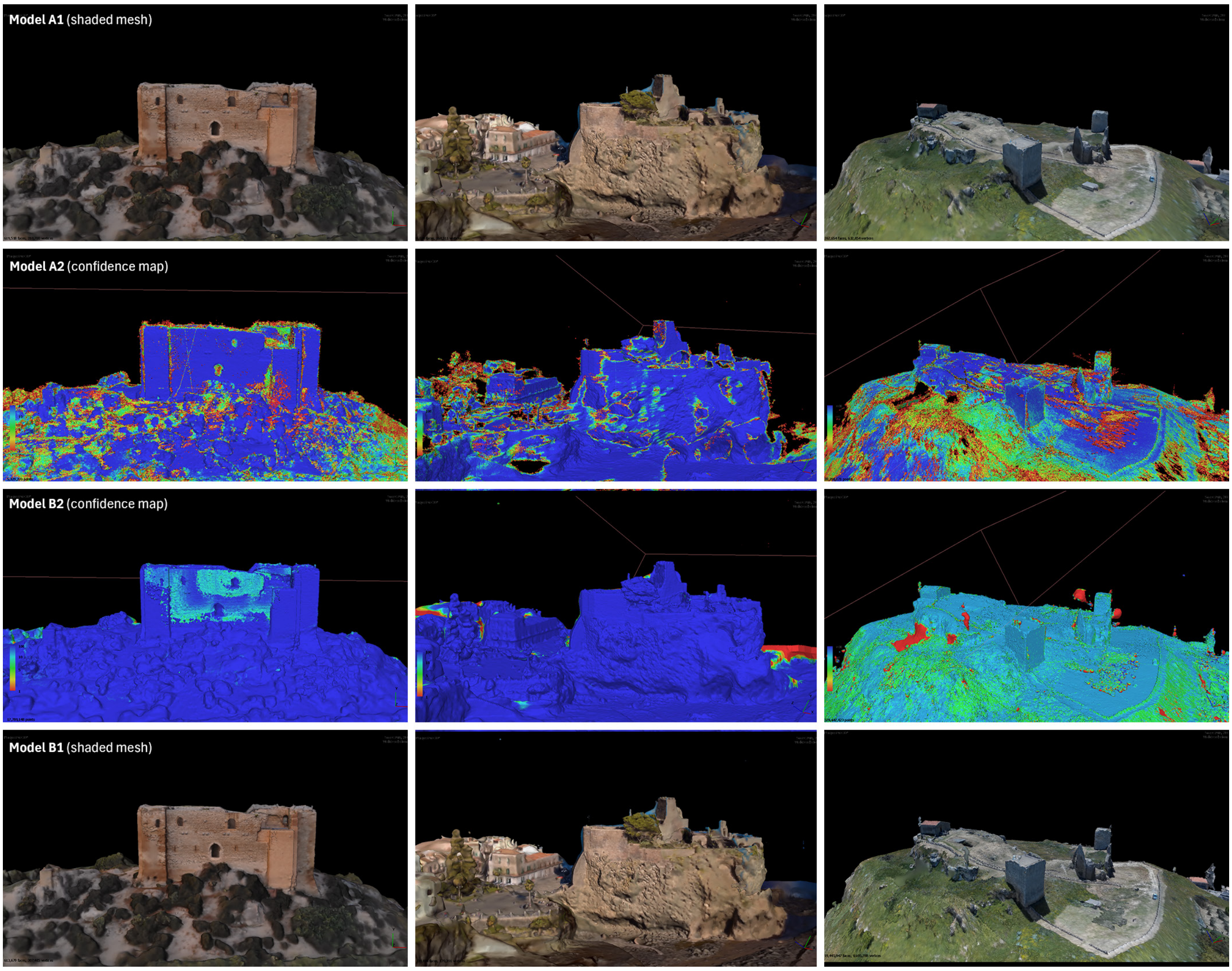

Figure 5. Visual comparison of first- and second-level 3D outputs: shaded mesh, confidence maps and reconstructed products.

and any irregularities of the source point cloud. This is useful when the aim is to preserve complex details, local surface variations, masonry discontinuities, edges, traces of decay or geometric irregularities. Mesh B1, generated from a very rich Dense Cloud, is more descriptive than the Mesh derived directly from Depth Maps, but the high number of faces may cause operational issues, with consequent difficulties in visualisation, texturing, export and management in external software, as occurred for the Castello di Agira. This behaviour confirms that greater geometric density does not necessarily coincide with greater operational efficiency of the model. The Point Cloud from Mesh (B2), by contrast, is obtained through geometric sampling of the surface of Mesh A1 and produces a point cloud from an already modelled surface. Its number of points may be very high, but this increase should not be interpreted as an increase in metric accuracy. Rather, it describes the sampling density of the Mesh and the possibility of reconverting a polygonal surface into point-based data useful for exports, visualisations or comparisons.

These distinctions make it possible to move from the procedural level to the quantitative comparison between the products obtained in the three case studies and to compare the different three-dimensional reconstruction paths, distinguishing four product types: Mesh from Depth Maps (A1), Dense Cloud from Depth Maps (A2), Mesh from Dense Cloud (B1) and Point Cloud from Mesh (B2). The comparison highlights two distinct behaviours, observable in the visual comparison of the reconstructed outputs (fig. 5) and in the quantitative data (tab. 3). In the Castelluccio di Gela and the Rocca di Aci Castello, the Mesh from Depth Maps (A1) and the Mesh from Dense Cloud (B1) return very similar values, indicating substantial coherence between the two reconstruction paths, although they start from different procedures. In the Castello di Agira, by contrast, the difference is marked: B1 is more than fifteen times denser than A1. This gap is consistent with the value of the Dense Cloud from Depth Maps (A2), which is also considerably denser at Agira than in the other two case studies: since B1 derives precisely from A2, a richer starting Dense Cloud is reflected in a proportionally heavier resulting Mesh, probably because of the larger initial dataset, the site's geometric complexity and the resulting computational complexity of the model. For the point clouds, in all three cases the Point Cloud from Mesh (B2) clearly exceeds the Dense Cloud from Depth Maps (A2), even by an order of magnitude. This higher value should not, however, be interpreted as an increase in metric accuracy, but as an effect of the sampling density of the already reconstructed Mesh surface: B2 derives from a resampling of the Mesh, whereas A2 derives directly from the dense photogrammetric process. The two clouds therefore have different origins and meanings, although they describe the same geometry through different densities and sampling logic. The evaluation of the products, however, cannot be reduced solely to the number of faces or points generated, but must consider the relationship between data origin, model weight, manageability and documentary purpose. In summary, the four outputs should not be read as equivalent products: A1 prioritises lightness and surface continuity, B1 preserves

Figure 6a. Rendered textured mesh models: overall views of Castelluccio di Gela, Rocca di Aci Castello and Castello di Agira.

Figure 6b. Shaded mesh models: for the three case studies detail views of the corresponding masonry surfaces.

greater geometric density, A2 maintains the point-based nature of the photogrammetric data and B2 responds mainly to export, comparison or visualisation needs in software environments requiring point-based data. The relevant issue for the research is not to establish which output is best in absolute terms, but to demonstrate that datasets derived from non-professional videos can generate differentiated, controllable models that can be selected according to documentary purpose. For the preliminary documentation of fortified heritage, this possibility is particularly useful: it enables a choice between lighter models intended for visualisation and denser models functional to surface reading, while keeping the entire process of generating three-dimensional data traceable.

Output	Castelluccio of Gela	Rocca of Aci Castello	Castello of Agira
A1 - Mesh from Depth Maps (faces)	619.518	342.282	1.262.654
A2 - Dense Cloud from Depth Maps (points)	5.329.209	2.528.687	20.296.171
B1 - Mesh from Dense Cloud (faces)	613.679	336.038	19.445.947
B2 - Dense Cloud from Mesh (points)	17.789.148	22.336.204	379.447.423

Comparison	Δ absolute	Δ%	Δ absolute	Δ%	Δ absolute	Δ%
Dense Cloud from Mesh vs Dense Cloud from Depth Maps	12.459.939	+233,80%	19.807.517	+783,31%	359.151.252	+1769,55%
Mesh from Dense Cloud vs Mesh from Depth Maps	-5.839	-0,94%	-6.244	-1,82%	18.183.293	+1440,09%

Table 3. Quantitative comparison of the first- and second-level 3D outputs generated for the three case studies.

5. Discussion and conclusions

The experimentation showed that non-professional video sequences, when subjected to a critical process of selection, filtering and camera optimisation (Gradual Selection), can be transformed into photogrammetric datasets usable for an initial three-dimensional record of fortified heritage (fig. 6). The main outcome does not lie in the production of a single model, but in the possibility of constructing controlled 3D data adaptable to different purposes of reading, visualisation and management.
Although the generated models cannot be considered certified metric products and present limitations in absolute georeferencing, the high density of frames provides considerable iconographic richness. This abundance of visual data makes it possible to obtain high-quality textures, characterised by visual continuity, surface detail and chromatic fidelity. Consequently, the main value of the methodology lies in the capacity of the texture to record information on the condition of surfaces and on

their evolution over time. The model thus becomes an effective tool for qualitative monitoring of decay, allowing the visual analysis of masonry surfaces, discontinuities, gaps, any crack patterns recognisable to the naked eye and vegetation growth. The proposed workflow, low-cost and applicable within limited timeframes, should be considered a preliminary and exploratory knowledge tool, not a procedure for producing certified metric deliverables. From this perspective, it is particularly useful in fragile, unstable or difficult-to-access contexts, often lacking systematic metric campaigns. The models obtained make it possible to frame the morphological relationships between masonry structures, rock and environmental context, offering a timely documentary basis for recognising situations that require subsequent checks. The effectiveness of the workflow also lies in its replicability: the methodology does not merely convert a video into frames, but defines a transparent, verifiable and controlled sequence of data selection, processing and evaluation. Under this condition, datasets derived from non-professional videos can acquire documentary value for the protection of cultural heritage, without being confused with planned metric surveys or specialist diagnostic investigations.

Transparency Statement

The authors declare that the artificial intelligence tools ChatGPT (OpenAI) and DeepL Write (DeepL) were used exclusively for minor language-editing purposes, including grammar, spelling and readability, and that all changes were reviewed and approved by the authors. No scientific content, analysis, results or interpretations were generated by AI.

References

Alsadik, B., 2016. Crowdsourced and web-published videos for 3D documentation of cultural heritage objects. *Journal of Cultural Heritage*, 21, 899–903. doi.org/10.1016/j.culher.2016.03.010.

Barazzetti, L., 2017. Network design in close-range photogrammetry with short baseline images. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, IV-2/W2, 17–23. doi.org/10.5194/isprs-annals-IV-2-W2-17-2017.

Bresc, H., Maurici, F., 2009. I castelli demaniali della Sicilia (secoli XIII-XV). In: Panero, F., Pinto, G. (Eds.), Castelli e fortezze nelle città italiane e nei centri minori italiani (secoli XIII-XV). Centro Internazionale di Ricerca sui Beni Culturali, Cherasco, 271–317.

Cardaci, A., Azzola, P., Versaci, A., 2026. The Rotonda of San Tomè in Almenno San Bartolomeo: rigorous digital 3D survey and data processing for a conscious and sustainable conservation. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-4/W19-2025, 37–44. doi.org/10.5194/isprs-archives-XLVIII-4-W19-2025-37-2026.

Cardaci, A., Versaci, A., 2019. Research and technological innovation for the knowledge, conservation and valorization of cultural heritage in Sicily. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W15, 247–254. doi.org/10.5194/isprs-archives-XLII-2-W15-247-2019.

Cardaci, A., Versaci, A., Azzola, P., 2024. 3D photogrammetric models from HDR images: the case study of the Fontanone Visconteo in Bergamo. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-2/W4-2024, 95–102. doi.org/10.5194/isprs-archives-XLVIII-2-W4-2024-95-2024.

Condorelli, F., Rinaudo, F., 2019. Benchmark of metric quality assessment in photogrammetric reconstruction for historical film footage. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* XLII-2/W11, 443–448. doi.org/10.5194/isprs-archives-XLII-2-W11-443-2019.

Fallavollita, F., Ugolini, A., 2017. New methodologies for the documentation of fortified architecture in the state of ruins. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-5/W1, 411–418. doi.org/10.5194/isprs-archives-XLII-5-W1-411-2017.

Federman, A., Santana Quintero, M., Kretz, S., Gregg, J., Lengies, M., Ouimet, C., Laliberte, J., 2017. UAV photogrammetric workflows: best practice guideline. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W5, 237–244. doi.org/10.5194/isprs-archives-XLII-2-W5-237-2017.

James, M.R., Robson, S., 2014. Mitigating systematic error in topographic models derived from UAV and ground-based image networks. *Earth Surface Processes and Landforms*, 39(10), 1413–1420. doi.org/10.1002/esp.3609.

Martin-Brualla, R., Gallup, D., Seitz, S.M., 2015. 3D time-lapse reconstruction from Internet photos. In: *Proceedings of the IEEE International Conference on Computer Vision*, ICCV 2015, 1332–1340.

Pavoni, G., Dellepiane, M., Callieri, M., Scopigno, R., 2016. Automatic selection of video frames for path regularization and 3D reconstruction. In: *Eurographics Workshop on Graphics and Cultural Heritage*, GCH 2016, 1–10. doi.org/10.2312/gch.20161376.

Schönberger, J.L., Frahm, J.-M., 2016. Structure-from-Motion revisited. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, 4104–4113. doi.org/10.1109/CVPR.2016.445.

Sun, Z., Zhang, Y., 2019. Accuracy evaluation of videogrammetry using a low-cost spherical camera for narrow architectural heritage: an observational study with variable baselines and blur filters. *Sensors*, 19(3), 496. doi.org/10.3390/s19030496.

Themistocleous, K., 2017. Model reconstruction for 3D visualization of cultural heritage sites using open data from social media: the case study of Soli, Cyprus. *Journal of Archaeological Science: Reports*, 14, 774–781.

Versaci, A., Fauzìa, L.R., Russo, M., Cardaci, A., 2020. The integrated fast survey for the risk assessment. A proposal for the safeguarding of the medieval castles in Central Sicily. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIV-M-1-2020, 893–900. doi.org/10.5194/isprs-archives-XLIV-M-1-2020-893-2020.

Westoby, M.J., Brasington, J., Glasser, N.F., Hambrey, M.J., Reynolds, J.M., 2012. Structure-from-Motion photogrammetry: a low-cost, effective tool for geoscience applications. *Geomorphology*, 179, 300–314. doi.org/10.1016/j.geomorph.2012.08.021.