

BENCHMARK OF METRIC QUALITY ASSESSMENT IN PHOTOGRAMMETRIC RECONSTRUCTION FOR HISTORICAL FILM FOOTAGE

F. Condorelli *, F. Rinaudo

DAD, Department of Architecture and Design, Politecnico di Torino, Lab G4CH – Laboratory of Geomatics for Cultural Heritage
(francesca.condorelli, fulvio.rinaudo)@polito.it

Commission II

KEY WORDS: Accuracy Requirement, Photogrammetry, 3D Reconstruction, Cultural Heritage, Historical Video Processing, Open Source Algorithms

ABSTRACT:

Quality assessment in photogrammetric processing is fundamental to obtain metric information and to reconstruct 3D models of Cultural Heritage, especially when it has been lost or changed over time. The determination of metric precision is technically challenging when dealing with historical films and videos that in many cases represent the only remaining traces of this heritage, which is useful for architectural, archaeological and restoration studies. This paper examines the suitability of existing photogrammetric software to evaluate the maximum possible metric accuracy for processing videos shot with fixed camera motions. In order to evaluate the metric quality obtained processing historical film footage with photogrammetric techniques, a benchmark was created on a new video dataset with the aim of reproducing the camera motions in which old video were shot. Three different camera motions were considered: Up/Down Motion-Tilting, Left/Right Motion-Trucking and Rolling Motion-Panning. The methodology was experimented on Valentino Castle in Turin, a monument inscribed in the UNESCO World Heritage List. Data were processed with the implementation of open source Structure-from-Motion algorithms and the results were analysed for the evaluation of metric quality. Results show the different maximum precision assessments according to the different typologies of camera motion. This research provides fundamental support to historical studies on Cultural Heritage, creating a sharing standard with zero-cost data and tools useful for both geomatics and restorers.

1. INTRODUCTION

Representing Cultural Heritage information and knowledge has received much attention in recent years due to a close collaboration between geomatics and restorers, which offer numerous benefits. An important aspect is the innovation in the documentation and conservation process resulting from the development of new geomatics technologies in support of historic and restoration studies. The increasing interest in causes and events that are destroying architectural and archaeological sites have heightened the need to create a basis of knowledge to transmit to future generations and effective strategies to manage and monitor heritage information. Rapid advances in the field of acquisition and processing of Cultural Heritage data have been attracting a lot of researchers and have led to a proliferation of studies in this direction. However, these rapid changes are presenting challenges to the professional partnership between those who recover data and those who use them, and this may cause difficulty in the collaboration. Of particular concern is the issue related to the accuracy of the metric survey. If this is not determined clearly by information providers, information users might not appreciate the importance and the limits of its possible uses (Tucci and Lerma, 2018). For this reason, it is important to establish guidelines on how to consider different accuracy and precision requirements in 3D surveys according to the different uses and purposes of the end-users. Especially in the Cultural Heritage field, it must be known that the metric survey is not the end of a project but the starting point of many applications and depends strongly on its metric quality. Geomatics experts play the important role of certifying the reliability of the survey and giving support to professionals who need metric information. Guaranteeing the quality of metric information extrapolated from 3D models

allows their correct use for restoration projects and monitoring applications.

The present research examines the metric quality assessment in photogrammetric processing and highlights its fundamental importance for 3D model reconstruction of Cultural Heritage. Accuracy requirements are necessary to extract metric information and obtain high-quality certified metric products, which is fundamental for documentation. Recent developments in the field of photogrammetry for Cultural Heritage documentation have led to rapid advances in the extraction of metric measurements, just in the cases where affordable metric information are really presents. The determination of metric precision is technically challenging when dealing with historical data from archives. This concerns particularly the processing historical films and videos, that is limited by the presence of characteristics that make it difficult to implement photogrammetry. Historical film footage stored in archives was not shot to be used for 3D reconstruction. In the majority of cases, it consists in movies, amateur videos or cinematographic reports. Another issue is that there is no way of knowing what kind of camera and film were used. In addition, footage is of low quality due to improper storage, which is a common problem for all historic material. The main disadvantage is the way in which the footage was shot. For this reason, processing and treating this data is a great challenge for photogrammetric research. However, they provide a huge amount of cultural and historic information and are of great value for architectural and restoration studies. For historic reconstruction they are sources of enormous informative potential because they testify the state of buildings, parts of a city and urban environment at a specific time, show famous heritage

monuments which were destroyed or damaged, and temporary architectures like those of international expositions. In fact, architectural and archaeological heritage has not always been documented by experts before it was lost, and in these cases, historical film footage represents the only source to reconstruct transformations over time. For this reason, many film libraries have understood the important role of the material they store and have recently started to share it with the public. The emerging needs to make preservation more efficient and to benefit from the use of Cultural Heritage material in archives have led research in the photogrammetric field to develop original applications in this direction.

The aim of this paper is to explore the relationship between historical information and the metric quality of historical data to support researchers and experts in historical research of Cultural Heritage. Therefore, this study presents a benchmark to evaluate the metric quality of results obtained from the processing of historical videos with photogrammetric techniques. The findings presented in this paper should make an important contribution to the field of representation and sharing of heritage information, since the creation of a benchmark on metric quality represents a standard that is useful both for geomatics experts, art historians and conservators.

This paper is divided into four parts. The first part deals with the identification of the type of camera motions used in the historical film footage and the state of the art in processing them. The second part is concerned with the methodology used for this study and the way in which the dataset was acquired. The photogrammetric workflow followed to process it is also described here. The third part presents the findings of the research, focusing on the evaluation of metric quality, and finally the fourth part analyses and discusses the results.

2. CAMERA MOTIONS AND RELATED WORKS

The authors developed a method in order to systematically assess the maximum precision reachable by implementing photogrammetric workflow to video frames shot with different types of camera motion for the creation of the benchmark. Despite the importance of this topic, researchers have not treated it in much detail and previous studies have not dealt with the determination of the metric quality evaluation of results from photogrammetric processing of historical film footage.

Recently the use of videos for the documentation of artefacts and heritage sites is becoming more and more common. Most research on video sequences have utilised to shoot ad-hoc cameras and a dense 3D reconstruction from videos has been proposed with the goal of having an accurate representation of the scene (Pavoni et al., 2016). Aerial video footage presents the same disadvantages of historical films, such as low resolution, blur-motion effect and redundancy of video frames, and for this reason could be compared. Previous studies explored the video frames usability for 3D modelling with the use of commercial software for Structure-from-Motion data processing (Cusicanqui et al., 2018). In the case of historical film footage the complete automation of software packages leads to no results and it is really necessary to control each step of the photogrammetric workflow.

In general, in historical film footage, it is very rare to find camera motions taken from multiple points of view of the

same object that create convergent views. If they are available, the application of bundle adjustment method allows the computation of all camera parameters and 3D object coordinates as well as the compensation of the systematic errors. Instead, it is much more common to find the following types of camera motions:

- 1) *Up/Down Motion* or *Tilting*: camera positioned in a fixed position and that takes the object by scrolling from top to bottom (or vice versa) in a vertical plane.
- 2) *Left/Right Motion* or *Trucking*: camera in motion along a fixed point and that takes the object by scrolling from right to left (or vice versa)
- 3) *Rolling Motion* or *Panning*: camera positioned in a fixed position and that takes the object by horizontally pivoting from right to left (or vice versa) on a central axis.

In these cases, the baseline between adjacent frames is absent or very small, for this reason the bundle adjustment could fail as the collinearity equations may be ill-conditioned or the rays cannot correctly intersect.

The following part presents the feasibility and limitations of processing film footage shot with these kind of camera motions and the analysis of the state of the art for each of them.

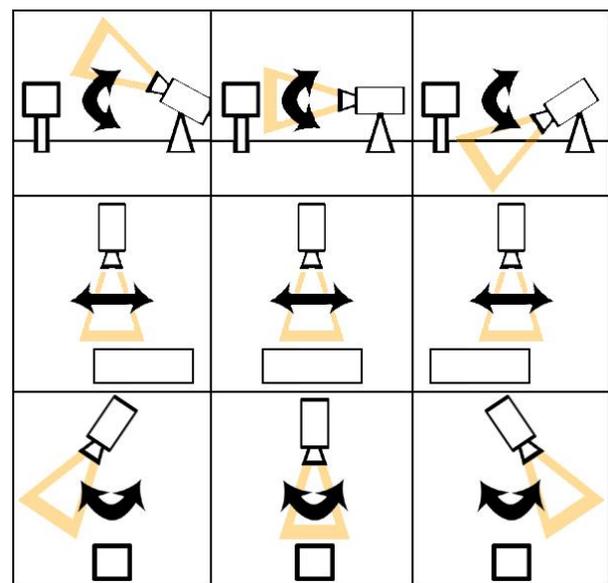


Figure 1. Scheme of the three types of camera motions: Up/Down Motion-Tilting, Left/Right Motion-Trucking and Rolling Motion-Panning.

2.1 Up/Down Motion-Tilting

In the case of the camera positioned in a fixed point and that takes the object by scrolling from top to bottom (or vice versa), the baseline between adjacent frames is absent and this could cause problems in the processing because the frames have too high overlap. A previous study examined the photogrammetric analysis of monocular video sequences without typical photogrammetric information for the recovering of camera parameters and the generation of 3D models. It demonstrated after a series of trial conducted on different datasets, that parameters for the image orientation and calibration were successfully obtained knowing the dimensions of some objects in the imaged scene, the pixel size and a perspective bundle adjustment (Remondino, 2004).

2.2 Left/Right Motion-Trucking

In the case of the camera positioned in front of the object and that takes it in motion by scrolling from right to left (or vice versa) there is more overlap between a frame and the other close. Also in this case, it is possible obtain the camera parameters of orientation and calibration.

In both cases the perspective collinearity model is simplified into a scaled orthographic projection (Remondino, 2004).

2.3 Rolling Motion-Panning

In the case of the camera positioned in a fixed point and that takes the object by rotating from right to left (or vice versa) there is absence of baseline and the conventional bundle method cannot solve the adjustment. A previous study demonstrated that the perspective camera model based on the conventional bundle method can be employed to calibrate rotating cameras that do not generate cocentric images. Otherwise a simplified camera model, that relates image correspondences only with a rotation matrix, can be used. The results obtained from the existing video do not respect the usual photogrammetric accuracy, mainly because of the very low image quality (Remondino and Börlin, 2004).

Therefore, this case could be related to the spherical photogrammetry theorised by Fangi (2007) and following studies. This is an analytical approach which works with a set of images taken from a unique point of view that produce a spherical panorama. It is obtained by stitching together several pictures which are then projected on a virtual sphere and later mapped in a plane with an equi-rectangular projection, using commercial software. If more than one panorama of the same scene, acquired from different point of view, are available, an appropriate orientation and the 3D reconstruction of the scene can be achieved (Barazzetti et al, 2010; Pisa et al., 2011). In these cases it has been demonstrated that it is possible reach good metric content with an average value for the error module between about 0.03 m and 0.015 m (D'Annibale et al., 2011).

3. VIDEO PROCESSING PIPELINE

Together, these studies indicate that, with some limitations, the feasibility of processing video sequences according to the three types of camera motions identified is demonstrated. Motivated by the disadvantages of previous researches and the absence of existing benchmark, a new dataset is presented with the aim of reproducing the way in which ancient video were shot.

A case-study approach was adopted to evaluate the effectiveness of the experimental methodology here presented. Data for this study were collected by authors in optimal conditions that is using a full frame calibrated camera, CANON EOS 5DS R, with fix focal length of 20 mm and known settings (focus, aperture, exposure). In the following table the specifications of the camera are shown.

Effective megapixels	50.60
Sensor size	36 x 24 mm
Sensor type	CMOS
Sensor resolution	8712 x 5808
Max. video resolution	1920x1080 (30p/25p/24p)

Table 1. Camera specifications.

3.1 Case study and dataset

Different videos of an existing historical building, the Valentino Castle in Turin, were shot. Valentino Castle is one of the “Residence of the Royal House of Savoy” and is included in the list of the UNESCO World Heritage since 1997. The present Valentino Palace of the Savoy dynasty derives from various planning phases which began in the mid-1500s. In conformity with the French *pavillon-système*, the architects Carlo and Amedeo di Castellamonte conceived the construction of an impressive building by doubling the existing architectural structure, enclosed by a pavilion roof and flanked by two tall, slender, lateral towers which are connected with terraced porticoes to two new pavilion roofs, towards Turin and linked by a semi-circular exedra. Following, conforming to the Eclectism culture spread the terraces linking the two towers were replaced by two big galleries. After many subsequent expansion and restoration works, the castle is now seat of the Politecnico di Torino.

The different camera motions of old films previously described were recreated. In particular principal façade, the two lateral wings and the courtyard of the castle were shot as following reported:

- 1) Up/Down Motion-Tilting: left wing of courtyard
- 2) Left/Right Motion-Trucking: façade, lateral wings and courtyard
- 3) Rolling Motion-Panning: façade and courtyard

3.2 Photogrammetric workflow with open source algorithms

Video frames were extracted from videos and processed with the software COLMAP, developed by ETH of Zurich, (available at <https://github.com/colmap/colmap>). COLMAP is an open-source Structure-from-Motion (SfM) and Multi-View Stereo (MVS) algorithm implementation and it was developed with the aim of design a general-purpose incremental SfM system for reconstruction of unordered photo collections. The use of open-source algorithms allows to control the quality of results in each step of the photogrammetric pipeline and avoids the blindly automatisms of commercial software packages. Moreover, current state-of-the-art SfM algorithms fail to register images with the problems previously highlighted in common with historical film footage and to produce fully satisfactory results in terms of completeness and robustness. The advantage of COLMAP is that significantly improved



Figure 2. A selection of the frames extracted from videos showing three camera motions: (1) Up/Down Motion - Tilting, (2) Left/Right Motion-Trucking, (3) Rolling Motion-Panning.

accuracy results while boosting efficiency at each step during the incremental reconstructions (Schönberger et al., 2016).

The results will become the benchmark for the evaluation of the quality of the historical video processing, and for this reason the settings and the workflow followed will be the same for the three different camera motions. In order to follow the same process for the three different cases, the same three steps of the COLMAP SfM sequential processing pipeline for the iterative reconstruction were followed: 1) Feature detection and extraction, 2) Feature matching and geometric verification, 3) Structure and motion reconstruction.

The software allows to set different reconstruction scenarios and in this case the best for to obtain high accuracy and efficiency is the *Video Sequences*, as a video presents consecutive frames with a too small baseline.

In the first step, feature detection and extraction find sparse feature points in the image and describes their appearance using a numerical descriptor. In the best case, like this used for the benchmark creation, the camera is calibrated, so it is possible manually specify intrinsic parameters. Generally in the case of historical film footage only partial or none EXIF information are available, but the software tries to find automatically camera and focal length information. The same camera took multiple pictures with the same lens and settings, so the same information may be shared between all the images. Then the intrinsic camera model must be chosen. In this case the intrinsic parameters are unknown a priori it is recommended to choose the *Simple Radial Camera Model* that is able to model distortion effects considering the following parameters: f , c_x , c_y , k_1 , k_2 , that is one focal length (f), two coordinates of the principal point (c_x , c_y) and

two radial distortion parameters (k_1 , k_2).

In the second step, feature matching and geometric verification finds correspondences between the feature points in different images. In the case study it was chosen the *Sequential Matching* mode developed for images acquired in sequential order by a video camera. In this case, consecutive frames have visual overlap and there is no need to match all image pairs exhaustively. For a better reconstruction the frame rate was reduced, it was increased the overlap and loop detection was enabled.

After the matching process, the incremental reconstruction process can start, and the results can be visualized in real-time. For better results the manually reconstruction was chosen and a *patch-match.cfg* file was written in which instructions for the reconstruction were given. In fact in these cases manually choosing the source images with most visual overlap leads to a better results as skipping some neighbours it is possible obtain larger baselines.

Data processed are stored in a customized database and could be easily managed. Finally the results obtained from the analysis and processing of the three types of camera motions are shown in Figure 3.

4. PRECISION METRIC QUALITY ASSESSMENT AND EVALUATION RESULTS

The analysis examined the results obtained from the previous processing step in order to assess precision of the models and to evaluate their metric quality. Only the Up/Down Motion-Tilting and the Left/Right Motion-Trucking cases were considered in this evaluation, whereas the Rolling Motion-Panning case refers to spherical photogrammetry, with the limit that only with more than one panorama it is possible to

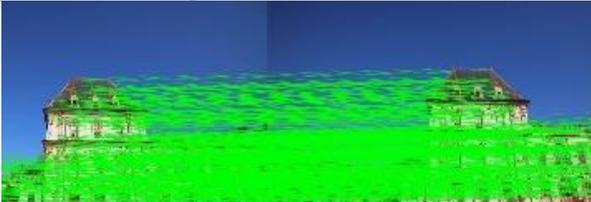
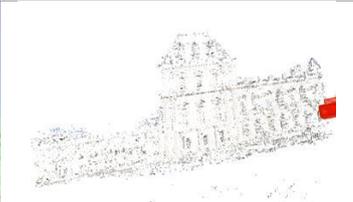
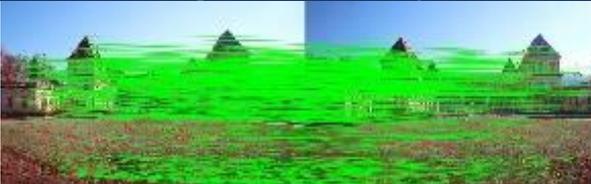
Feature detection and extraction	Feature matching and geometric verification	Structure and motion reconstruction
		
Case 1: Left wing of courtyard	Camera Motion: Up/Down Motion-Tilting	Number of video frames: 83
		
Case 2: Façade and courtyard	Camera Motion: Left/Right Motion-Trucking	Number of video frames: 98
		
Case 3: Façade and courtyard	Camera Motion: Rolling Motion-Panning	Number of video frames: 24

Figure 3. Video frame processing: the three steps of the COLMAP SfM sequential processing pipeline for the iterative reconstruction.

reach a good metric quality, as previously showed.

4.1 Standard deviation of Final Cost

For the precision analysis, the values of the Final Cost from the bundle adjustment report of the process were examined. Final Cost represents the average of the reprojection error over all image observations and it is expressed in pixel. All values of Final Cost for each case were used for the calculation of the Mean and the Standard Deviation and reported in the following graphs to analyse the trend of the data.

Camera motion	Mean	Standard Deviation
Up/Down Motion	0.36 [px]	0.10 [px]
Left/Right Motion	0.47 [px]	0.13 [px]

Table 2. Mean and Standard Deviation value of Final Cost in pixel.

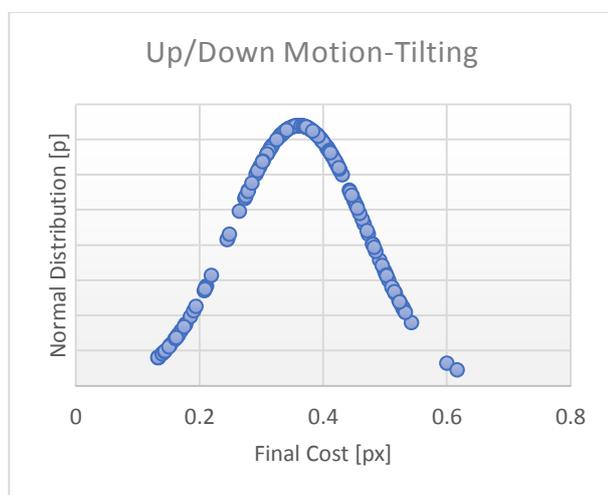


Figure 4. Normal Distribution of Final Cost Value for Up/Down Motion-Tilting.

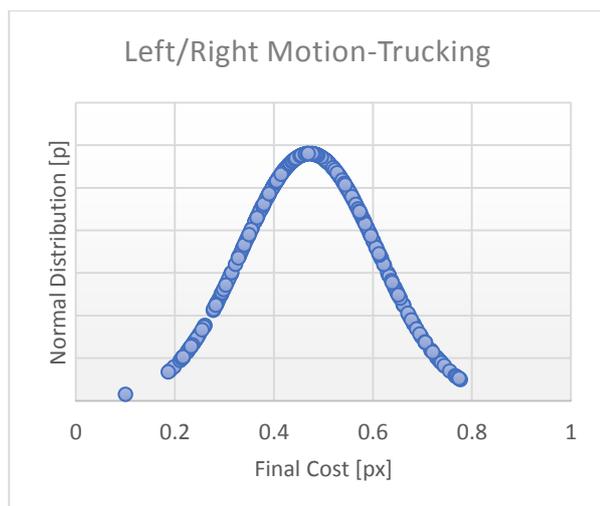


Figure 5. Normal Distribution of Final Cost Value for Left/Right Motion-Trucking.

Moreover, the minimum and the maximum values were highlighted and transformed in centimetre with the Ground

Sample Distance (GSD) calculation. The results are set out in Table 3,4 and 5.

Camera motion	Min Final Cost	Max Final Cost
Up/Down Motion	0.13 [px]	0.60 [px]
Left/Right Motion	0.10 [px]	0.77 [px]

Table 3. Min and Max value of Final Cost in pixel.

Camera motion	Distance	GSD
Up/Down Motion	15 [m]	1.2 [cm/px]
Left/Right Motion	120 [m]	11.2 [cm/px]

Table 4. GSD Calculation.

Camera motion	Min Final Cost	Max Final Cost
Up/Down Motion	0.1 [cm]	0.8 [cm]
Left/Right Motion	1.1 [cm]	8.7 [cm]

Table 5. Min and Max value of Final Cost in centimetre.

Interestingly, from graphs comes out that in both cases the trend of the curves is similar to a Gaussian Distribution.

It is apparent from the tables that, comparing the two results, it can be seen that the first case has reached a major metric quality than the second one. It is caused by the fact that the videos were shot at different distances, and it could affect the results.

However, both cases are suitable for high-quality photogrammetric reconstructions. In fact, for the property of Mean (μ) and Standard Deviation (σ), the 99,73% of values of Final Cost are included in the range of $\mu \pm 3\sigma$, that in this case is of the order of 1 pixel. Assuming this value as maximum error and considering the distances of 15 m and 120 m, the maximum performance in terms of precision is 10^{-3} of the taking distance.

Considering these findings, the present study raises the possibility to use them as benchmark of metric quality for the future video processing and represent the maximum level of metric quality reachable by the analysed specific camera motions.

5. CONCLUSIONS

The experimental work presented here is one of the first investigations into the level of quality of results it is possible to reach when processing historical film footage. The study develops an innovative methodology for investigating and evaluating the precision of the photogrammetric reconstruction of Cultural Heritage from historical videos, according to different camera motions. The findings suggest that the pipeline followed to process videos lead to a high-quality result of metric precision assessment that will be used as a benchmark for the next stage of video processing. These results represent, in fact, the maximum performance of metric precision. Future research should further validate the methodology experimented here and continue the metric evaluation studying ways to test relative and absolute metric accuracy of point clouds with different level of density.

This study has a number of important implications for future practice especially in the field of Cultural Heritage. The virtual reconstruction of lost or damaged Cultural Heritage that appears in historical film footage allows historians and architects to explore the cities as they were in the past and to understand their evolution and the previous state of buildings and urban environments. Furthermore, this study shows that experimenting new methodology for documenting and managing historical data can be a good approach to help geomatics and restorers exchange skills and create common and shared standards.

The just started study will also allow the revisiting of the videogrammetry technology proposed in the past and that now, thanks to the technological advances, could be rethink to give another possible metric survey technology to be used inside an integrated survey of complex architectural and natural assets.

ACKNOWLEDGEMENTS

This work was supported and funded by the GAMHer project (Geomatics data Acquisition and Management for landscape and built Heritage in a European perspective), a 3-year project financed under the Italian PRIN 2015 framework (Progetti di Ricerca di Rilevante Interesse Nazionale).

REFERENCES

- Barazzetti, L., Fangi, G., Remondino, F., Scaioni, M., 2010. Automation in multi-image spherical photogrammetry for 3D architectural reconstructions. Proc. of 11th Int. Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST 2010), Paris, France. doi.org/10.2312/PE/VAST/VAST10S/075-081.
- Caraceni, S., Carpenè, M., D'Antonio, M., Fiameni, G., Guidazzoli, A., Imboden, S., Liguori, M. C., Montanari, M., Trotta, G., Scipione, G., Hanegreefs, D., 2017. I-media-cities, a searchable platform on moving images with automatic and manual annotations. 23rd International Conference on Virtual System & Multimedia (VSMM), Dublin, Ireland. doi.org/10.1109/VSMM.2017.8346274.
- Charvillat, V., Tonazzini, A., Van Gool, L., Nikolaidis, N., 2009. Image and Video Processing for Cultural Heritage. Hindawi Publishing Corporation, EURASIP Journal on Image and Video Processing, Volume 2009. doi.org/10.1155/2009/163064.
- Condorelli, F. and Rinaudo, F., 2018. Cultural Heritage reconstruction from historical photographs and videos. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2, 259-265. doi.org/10.5194/isprs-archives-XLII-2-259-2018.
- Cusicanqui, J., Kerle, N., Nex, F., 2018. Usability of aerial video footage for 3-D scene reconstruction and structural damage assessment. Nat. Hazards Earth Syst. Sci., 18, 1583-1598. doi.org/10.5194/nhess-18-1583-2018.
- Dameri, A., 2009. Storia e architettura. Il castello del valentino. Le residenze sabaude, Allemandi Torino, 108-122.
- D'Annibale, E., Piermattei, L., Fangi, G., 2011. Spherical photogrammetry as emergency photogrammetry. XXIIIrd International CIPA Symposium, Prague, Czech Republic. ISBN: 978-80-01-04856.
- Fangi, G., 2007. The multi-image spherical panoramas as a tool for architectural survey. XXI International CIPA Symposium, Athens, Greece.
- Pavoni, G., Dellepiane, M., Callieri, M., Scopigno, R., 2016. Automatic selection of video frames for path regularization and 3D Reconstruction. EUROGRAPHICS Workshop on Graphics and Cultural Heritage. doi.org/10.2312/gch.20161376.
- Pisa, C., Zeppa, F., Fangi, G., 2011. Spherical photogrammetry for Cultural Heritage—san Galgano Abbey and the Roman Theater, Sabratha. ACM J. Comput. Cult. Herit. 4, 3, Article 9.
- Remondino, F., 2004. Character reconstruction and animation from monocular sequence of images. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXV-B5,702-707. XXth ISPRS Congress, Istanbul, Turkey.
- Remondino, F., Börlin, N., 2004. Photogrammetric calibration of sequences acquired with a rotating camera. ISPRS Archives, Volume XXXIV-5/W16.
- Schönberger, J. L., Frahm, J. M., 2016. Structure-from-motion revisited. IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Vol. 2016, 4104-4113, IEEE Computer Society.
- Schönberger J.L., Zheng, E., Frahm, J.M., Pollefeys, M., 2016. Pixelwise View Selection for Unstructured Multi-View Stereo. ECCV 2016, Lecture Notes in Computer Science, vol 9907, Springer, Cham.
- Schöps, T., Schönberger, J. L., Galliani, S., Sattler, T., Schindler, K., Pollefeys, M., Geiger, A., 2016. A Multi-View Stereo Benchmark with High-Resolution Image and Multi-Camera Videos. 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), Honolulu, HI, USA.
- Tucci, G., Lerma, J.S., 2018. Special Issue GEORES2017. Geomatics and Restoration: Conservation of Cultural Heritage in the Digital Era. Applied Geomatics, Volume 10, Issue 4, 277–278.
- XiaoWen, W., Jon Mills, X., Guidi, G., González-Aguilera, D., 2018. Geoinformatics for the conservation and promotion of Cultural Heritage in support of the UN Sustainable Development Goals. ISPRS Journal of Photogrammetry and Remote Sensing. doi.org/10.1016/j.isprsjprs.2018.01.001, 2018.