

NEW IMAGING MOBILE MAPPING DEVICE BASED ON HIGH RESOLUTION VIDEOGRAMMETRY FOR LARGE-SCALE OUTDOOR 3D RECONSTRUCTION

P.Ortiz-Coder¹, R.Cabecera^{1*}

¹ 2Freedom Imaging Software and Hardware S.L. Calle Serranía de los Caballeros, Portal 1, Local, 06800 Mérida, Badajoz, Spain –
info@2freedom.net, rcs@aiuestudio.com

KEY WORDS: Videogrammetry, Mobile Mapping System, Visual SLAM, Dense 3D Reconstruction..

ABSTRACT:

In recent years, a new generation of instruments has appeared that are motion-based capture. These systems are based on a combination of techniques, among which LIDAR stands out. In this article we present a new proposal for a 3D model generation instrument based on videogrammetry. The prototype designed consists of two cameras connected to a computer system. One of the cameras is in charge of running VisualSLAM and guiding the user in real time at the moment of data acquisition; the other camera, with a higher resolution, saves the images and, thanks to a refined 3D-Based frame selection algorithm, processes them using automatic photogrammetric procedures, generating one or more point-clouds that are integrated to give way to a high-density and high-precision 3D colour point-cloud.

The paper evaluates the proposal with four case studies: two of an urban nature and two related to historical heritage. The resulting models are confronted with the Faro Focus3D X330 laser scanner, classic photogrammetric procedures with reflex camera and Agisoft metashape software and are also confronted with precision points measured with a total station. The case studies show that the proposed system has a high capture speed, and that the accuracy of the models can be competitive in many areas of professional surveying and can be a viable alternative for the creation of instruments based on videogrammetry.

1. MANUSCRIPT

Acquiring data in a simple way, while the user is walking, and then generating a reliable and accurate point cloud, is a growing need in the field of engineering, heritage and architecture. If we focus on terrestrial 3D capture sensors over the last two decades, active sensors have taken a key role in providing solutions to these kinds of needs. The 3d laser scanner technology in all its variants allows the user to obtain 3d models of millimetric precision. This type of static laser scanners are more accurate, but slower to capture than mobile laser scanners, which allow faster data acquisition but, on the other hand, are less accurate (Zlot et al. 2014). Latest generation handheld scanners, *e.g. Zeb Horizon by Geoslam or BLK2GO by Leica Geosystem* allow capturing while the user is on foot, being optimal for indoor and outdoor use in different distance ranges, which can vary from 10m-50m while maintaining centimetric accuracies.

Image-based modelling is currently an alternative to 3d active sensor measurement (Remondino et al. 2017). But photogrammetry has some limitations when it comes to automatic capture with handheld devices: photographs must be taken by hand one by one, camera settings need to be adjusted every time light conditions change, capturing pavements or floors requires a pole and remote triggering or similar technology, and sometimes capturing can become very tedious if the area to be documented is large, requiring thousands of photographs, etc. Although in specific cases terrestrial photogrammetry is an ideal solution (Cerrillo-Cuenca, Ortiz-Coder, y Martínez-del-Pozo 2014; James, Robson, y Smith 2017), especially in environments with heterogeneous textures with a limited surface area, it seems clear that in certain cases

terrestrial photogrammetry is no longer operational for the reasons explained above.

Videogrammetry is also an alternative for documenting long scenes (Schöps et al. 2017) and is currently being investigated by the computational vision community, where one of the target applications is the documentation of environments for autonomous cars (Barua et al. 2019). In computational vision, the main objective is to generate 3D models quickly, usually in real time, with accuracy as a secondary priority (Pollefeys et al. 2008; Schöps et al. 2017).

Videogrammetric capture can be done in a monocular way (Dorian Gálvez-López, Marta salas, y Juan D tardós s. f.), with stereocameras or with multicameras. And, in parallel, videogrammetry may have, as a priority, to obtain measurements and 3D models in real time (Pollefeys et al. 2008), or, as is the case in this article, to obtain millimetric or centimetric accuracies that can be transformed into an alternative or complement to measuring instruments in professional works (Luhmann et al. 2006).

Regarding the acquisition of videogrammetric data with monocular cameras and with priority in obtaining precision measurements, we can mention different projects as the product MX7 by Trimble Inc. where the applications are, mainly, to make point-to-point or manual measurements, through georeferenced images, in this case with GNSS.

This paper is focused on obtaining accurate 3D models through videogrammetry in large scale environments and with a handheld type of capture, i.e., the user carries the device and performs the data acquisition as moves around. The objective of obtaining maximum accuracy and resolution in the results

* Corresponding author

through this form of data acquisition has some basic challenges: on one hand a high-resolution video is needed and a global shutter type sensor is required to avoid deformations with camera movement, especially with a high-resolution sensor and unfavourable lighting conditions. On the other hand, the user, while capturing the video, does not know if he/she has made any rapid movement that could have caused the continuity of the trajectory to be lost or if he/she has made any other movement with the camera that cannot be resolved later with photogrammetry (for example, camera autorotations without displacement of the centre of projection) (Luhmann et al. 2006). And finally, the frame selection must be optimal to achieve the best results, avoiding excesses or defects in the number of images that may cause inaccuracies or, simply, a lack of sufficient connection between the images causing a break in the formation of the trajectory, considering the form of continuous capture that the video has (Torresani y Remondino 2019).

In this paper we present a new proposal for data acquisition using a handheld device with two cameras, one of low resolution that allows real-time navigation and assists the user in their movements, and a second camera with high resolution and image quality, which, through an algorithm designed for this purpose, selects the best frames to perform the photogrammetric process, also designed for this purpose. In order to evaluate the proposed system, four case studies are presented in which the resulting models have been compared with point clouds generated by the Faro Focus3D X330 laser scanner and by photogrammetric software available on the market, in addition to having calculated the errors through precision points measured with a total station. The case studies have been carried out in urban and historical heritage environments.

It is also a challenge to offer solutions that allow the use of the system proposed in this article to non-specialists in the sector, universalising the use of videogrammetry and allowing access to this information to professionals from other fields.

This article also shows the evolution in the design of this device from its conception in 2019 to the present day, (figure 1).

2. SYSTEM OVERVIEW

The designed device aims to capture the environment while the user moves around the scenario. To do this, the user must carry the device in his hand and focus it to the area to be captured while moving; in turn, the device will be connected to a computer system that will manage the cameras, store the frames, and run software to guide the user in real time. Subsequently, the optimal frames will be selected and the photogrammetric process will be carried out with the selected frames and the capture data in real time.

The operation of the proposed device uses two cameras for two different functions: camera B is used for high-resolution image capture and camera A is used for real-time application of a localization system.

To ensure the highest quality image acquisition, an industrial camera (camera B) of JAI Go-5000C brand with a resolution of 2560 x 2048 pixels and a 1" global shutter sensor has been chosen; this way we avoid deformation of the images, as the user carries the device in his hand while walking. The framerate used can vary from 2fps to 12fps, depending on the use and application. The field of view is (H x V) 89° x 76,2°, the lens

used is an Azure 6.5mm focal lens, although this can be interchangeable with others if the scenario requires it.

Camera A is from ELP brand, has a field of view of (H x V) 170° x 126°, a framerate of 25fps and a resolution of 640x480pix (VGA). Cameras A and B were calibrated internally using a calibration standard (Zhang 2000). The external calibration between both cameras was also calculated by simultaneously capturing two walls at an angle of 90° with 15 targets measured with the TOPCON Robotic total station with 1" accuracy. Through a seven-parameter transformation (Luhmann et al. 2006) and considering the previously calculated internal parameters, rotation and translation were performed for each camera in the same coordinate system. Cameras A and B are encapsulated in a housing, with a baseline of 4 cm, and connected via USB to a computing device (see figure 1).

Camera A, through the ORBSLAM application (Raúl Mur-Artal, J. M. M. Montiel, y Juan D. Tardós 2017) calculates the main keyframes that will serve as a starting point for the refined image selection system. ORBSLAM performs a real-time "bundle adjustment" that together with a loop closure function, establishes a keyframe detection system based on the geometry of the scenario, called 3D-Based (Torresani y Remondino 2019). It must be considered that cameras A and B are different and therefore, the images selected by camera A do not have to match for camera B images, for that reason a new "refined" image selection process has been designed (Ortiz-Coder y Sánchez-Ríos 2019; Ortiz-Coder y Sánchez-Ríos 2020).

The system uses camera A to calculate keyframes in real time, but also to help the user not to make incorrect movements, because if the user is too fast or is interrupted by objects in front of the camera or tries to make certain movements that the bundle adjustment cannot solve, the tracking will be lost and the system prompts the user to find the last point just before getting lost.

The VSLAM and the initial keyframe selection system is an on-line system, but the image processing through the refined keyframe selection system and the subsequent photogrammetric processing is an off-line process, which will demand a computational cost that will also be evaluated later in the results section.

Considering that the device needs to be connected to a computing system and must be transported by the user while walking, several prototypes have evolved to a final cylindrical design that can be adapted to a pole, can be carried with a handle or directly by hand (see figure 1).



Figure 1. Evolution of the design of the prototypes. First prototype while scanning the Roman aqueduct of "Los Milagros" (above). Last version of the prototype optimising the design and ergonomics by carrying it on a handle (bottom left) and carrying it directly by hand (bottom right).

Once the images from camera B have been captured and stored and the keyframes have been calculated through the visualSLAM of camera A, we proceed to find the coincidences in time of capture of the keyframes with camera B. This is done through the synchronisation of both cameras. Afterwards, the frames of camera B that have not suffered a displacement of their projection centres higher than a tolerance value D_i are eliminated; in this way we eliminate the frames that belong to those moments in which the user was standing still. Next, we determine whether there are rotations of a set of images that is higher than a tolerance level δ_i and, if so, we check whether these frames have undergone less displacement than a value D_i ; in this case, we understand that these cameras could not be solved in a bundle adjustment with precision guarantees. Therefore, the algorithm designed eliminates these images.

Then, the optimal connection between the selected frames is checked by calculating the number of homologous points between consecutive frames. If the number of homologous points is less than a minimum tolerance value p_i , n intermediate frames are included between these images. Subsequently, we recalculate whether any disconnection still exists and if the number of homologous points between consecutive images is still less than p_i , we create a new group of independent images from the disconnection zone. We will call this group of connected images a segment.

Relative orientation of the images contained in each segment is performed independently. The relative orientation is performed through a direct solution and a subsequent bundle adjustment (Rupnik, Daakir, y Pierrot Deseilligny 2017) until the divergences are minimised through the different iterations. In

this phase 3D points are computed from the tie points orientation results. Then, a minimum squares adjustment of the camera positions obtained for camera A is performed on the trajectory calculated in the VSLAM procedure for camera B (Mikhail, Bethel, y McGlone 2001). This operation will be performed independently for each segment, so that all the images of the different segments are in the same coordinate system, even without scale.

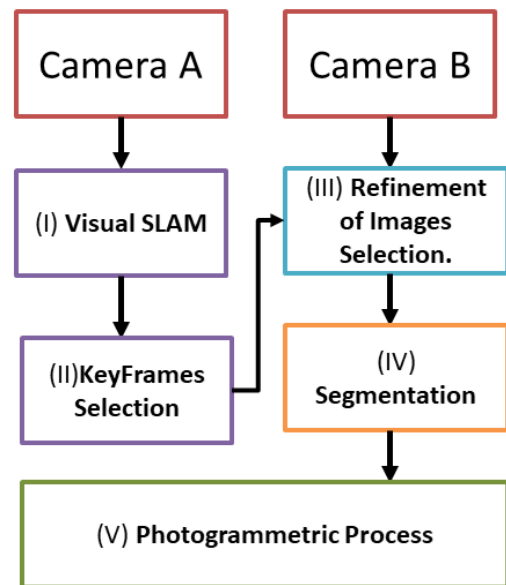


Figure 2. Refined image selection algorithm.

For the 3D reconstruction using photogrammetry we used the open source software MICMAC (Rupnik et al. 2017) belonging to the French National Mapping Agency (IGN) and the National School of Geographic Sciences (ENSG). Using this software, depth maps and coloured point clouds are calculated through the orientations made on each segment of the previous phase.

Data acquisition with this device, in a lineal mode, naturally generates an accumulation of errors proportional to the displacement. To minimise this drift error, the user is advised to make loops in the capture, capturing previously captured areas, or to place targets along the path with associated external coordinates that become Ground Control Points that can be used in the relative orientation as fixed elements, minimising drift errors and scaling the result.

3. RESULTS

In order to evaluate the results of the proposed system, four case studies were carried out in which data acquisition times, processing times, the resolution of the resulting point clouds, the errors made were measured and compared, depending on the case study, with the FARO Focus3D X330 laser scanner and with classic photogrammetric procedures, as well as with precision point measurements taken with a total station.

3.1. Case Study 1.

The first case study is a building in the city of Mérida (Spain). This building is 70 x 20 x 17 meters high. The data acquisition with the proposed system was performed at a distance from the façade varying between 13m and 20m, due to the conditions of the surrounding streets. The acquisition was carried out on foot, going around the building and ending at the same place where it

started. The duration of the capture was 5min and 5sec. The process of image selection and photogrammetric processing, as defined in the methodology section, took 1h 25min and a cloud of 49 million points was obtained.

To compare the point-cloud, a 3D scan was performed with the FARO Focus3D X330 laser scanner, where it was placed in 10 different positions at a resolution of 9mm at a distance of 15m, and a registration was performed with the Faro Scene software, which yielded a residual error of the standard deviations of 95mm. The resulting point-cloud from the laser scanner has 17,112,258 points.

The comparison between the point clouds obtained with the proposed system and the laser scanner was performed using CloudCompare 2.11.3 software (Anoia) and the results can be seen in figure 3. The mean distance between both models was 91mm and the standard deviation was 0.3.

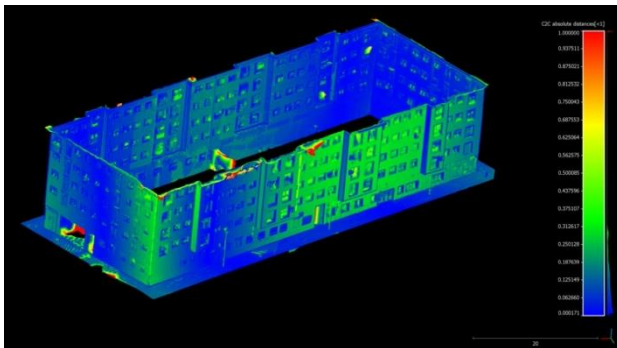


Figure 3. Comparison with laser scanner Faro Focus3D. Distance map between both 3D models resulted.

3.2. Case Study 2.

In order to know the accuracy in smaller urban captures, where the drift error does not affect, we proceeded to capture a façade in the city of Mérida (Spain) with dimensions of 20.6m long by 8.90m average height. The data acquisition with the proposed system was performed at a distance of 8 metres from the façade and with a single pass, in a single direction, carried out in 7 seconds a point-cloud of 2,624,325 points was generated. To compare the data, we used the ground data captured with the Faro Focus3D X330 laser scanner with a resolution of 2mm at a distance of 5m. The resulting point-cloud was 4,977,582 points. To compare both point-clouds, the CloudCompare 2.11.3 software (Anoia) was used to obtain a map of the minimum distances between the two models, where the average distance is 9mm and the standard deviation is 0.095.

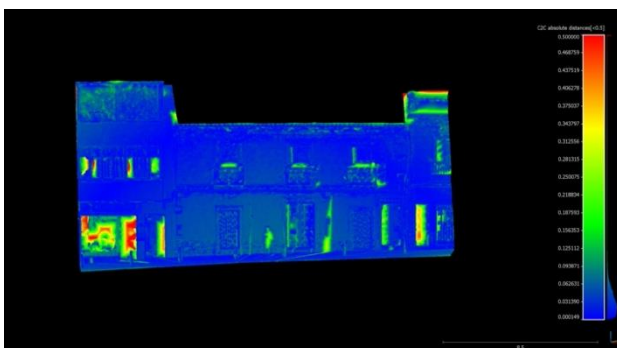


Figure 4. Comparison with LS. Distance map between both 3D models(Top) and photorealistic model with the proposed system(Botton).

3.3. Case Study 3.

The third case study is the 3D digitisation of a part of a Roman aqueduct in the city of Mérida, called "Los Milagros" (1st century AD). In this case study we wanted to evaluate the behaviour of the device when digitising an architectural monument, which is a World Heritage Site (UNESCO 1993) and at the same time evaluate the accuracy of the system through the distance of the sensor to the monument and, on the other hand, make a comparison with the traditional photogrammetric method by taking the photographs manually and using commercial software for data processing.

To carry out the evaluation with the proposed system, three different point-clouds have been generated, captured at three different distances from the section of the monument selected to be digitised, about 60m long and about 23m high. The first trajectory was taken at 5m from the monument, the second at 12m and the third at 20m. On the other hand, 60 points equally distributed on the two sides of the documented aqueduct section were measured with a Pentax V-227N prismless total station (Pentax Ricoh Imaging Company, Ltd, Tokyo, Japan) with an angular accuracy of 7' (ISO 17123-3:2001) and a distance measurement error of $3 \text{ mm} \pm 2 \text{ ppm}$ (ISO 17123-3:2001).

The point cloud resulting from the proposed procedure was scaled and referenced using 5 points measured with the total station. These points were not used for the precision measurement, in which the Euclidean average distance error (δ_{avg}) and the RMSE value for the three components X, Y and Z were calculated.

In parallel, photographs were taken with the Canon EOS1300D camera with the EFS 18-55mm lens, although we only used the 18mm focal length to take the photographs. As with the proposed device, three capture sessions were carried out at the three proposed distances: at 5 metres, 35 shots were taken, at 12 metres, 41 images were captured and at 20 metres, 43 images were captured. The resolution of the camera was set at 2592 pixels \times 1728 pixels to allow for a fair comparison. The overlap used was 80% in all shots. The software for image processing was Agisoft Metashape 1.5.4 by the company Agisoft LLC, located in St. Petersburg, Russia.



Figure 5. Comparison with Agisoft Metashape and Photorealistic model with the proposed system.

The three resulting point clouds obtained with Agisoft Metashape were also evaluated in an analogous way to the models obtained with the proposed device, obtaining average error and RMSE values for the three components X, Y and Z.

Table 1 shows the results of the errors measured for the different distances and for the two methods evaluated: the proposed system and the methodology of manual image acquisition with the camera and its processing with Agisoft Metashape. Below is a table showing the evolution of the errors according to the distance for both methodologies.

Prototype Proposed + Algorithm					Canon 1300D + Agisoft Metashape			
Distance 5m								
	Error Vector X(mm)	Error Vector Y(mm)	Error Vector Z(mm)	Error (mm)	Error Vector X(mm)	Error Vector Y(mm)	Error Vector Z(mm)	Error (mm)
Average Error				12				11
RMSE	5	8	8	8	4	9	8	12
Distance 12m								
	Error Vector X(mm)	Error Vector Y(mm)	Error Vector Z(mm)	Error (mm)	Error Vector X(mm)	Error Vector Y(mm)	Error Vector Z(mm)	Error (mm)
Average Error				26				23
RMSE	21	16	10	16	12	18	17	28
Distance 20m								
	Error Vector X(mm)	Error Vector Y(mm)	Error Vector Z(mm)	Error (mm)	Error Vector X(mm)	Error Vector Y(mm)	Error Vector Z(mm)	Error (mm)
Average Error				46				35
RMSE	32	30	38	33	18	24	24	39

Table 1. Comparison of errors between the proposed system and the classical photogrammetry methodology, measured with control points.

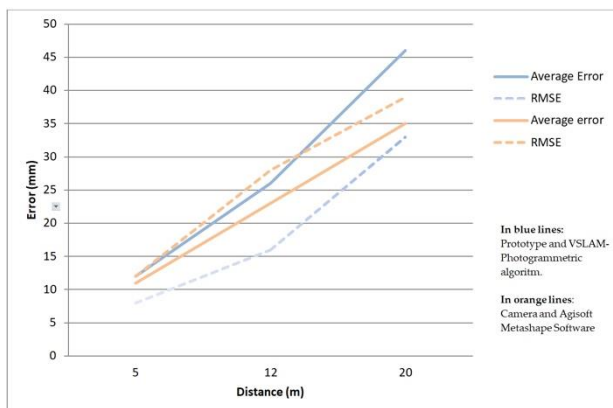


Figure 6. Graph showing the comparative errors of the proposed system and the classical methodology through the distance from the sensor to the monument.

3.4. Case Study 4.

In order to check the proposed system on archaeological sites and to know the error of the results with high accuracy, an experimental test was carried out on the remains of five rooms with mosaic floors in three of them, located in the archaeological site of the Casa del Mitreo (dated between the 1st and 2nd century BC), in the city of Mérida.



Figure 7. Archaeological remains of the rooms and mosaics in the monumental complex of the "Casa del Mitreo", in the city of Mérida (Spain).

First, 40 targets distributed throughout the different rooms were measured with the Pentax V-227N total station and then the data acquisition was carried out with the proposed system by performing a trajectory around the remains until the entire scene was documented. In this case, the trajectory ended at the starting point in order to compensate for drift errors. Finally, the FARO Focus3D X330 laser scanner was positioned in 11 strategically chosen locations so that the entire scene was well documented, and different spheres were also placed in order to minimise the alignment error.

The time used for data acquisition with the proposed system was 12 minutes, while the time used for scanning the entire scenario with the LS was 220 minutes, since images were also captured with the laser scanner. Regarding the data processing with the proposed system, it was carried out in a fully automatic way, taking 690 minutes. The laser scanner data processing was carried out using the FARO Scene software; 115 minutes of user time and automatic processing (without human intervention) were spent on the registration of the scans and the application of textures.

The resolution achieved by the proposed system in this case study is 1.4mm mean point-to-point distance, while the LS is 0.5mm.

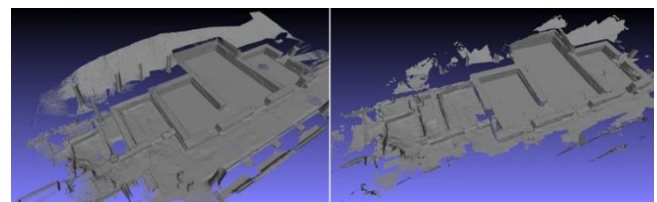


Figure 8. Visual comparison of 3D models generated with the laser scanner (left) and with the proposed videogrammetric system (right).

Prototype Proposed + Algorithm					Faro Focus3D X330			
	Error Vector X(mm)	Error Vector Y(mm)	Error Vector Z(mm)	Error (mm)	Error Vector X(mm)	Error Vector Y(mm)	Error Vector Z(mm)	Error (mm)
Average Error				8				11
RMSE	6	4	5	5	5	8	7	6

Table 2. Comparison of errors between the systems compared: proposed system and the Faro Focus3D X330 laser scanner.

4. CONCLUSIONS

In this paper we present a new proposal for a handheld 3D measurement instrument based on videogrammetry which, through its user-guided capture system and its image selection algorithm, allows complex trajectories and a longer duration, being able to capture larger scenarios.

The photogrammetric procedure performed with the high-resolution images facilitates the achievement of millimeter accuracies at short ranges (less than 1.5m from the object to the sensor) and of a few centimetres at distance ranges of 15-25m. Capture is 17 times faster than the laser scanner used and the type of user interaction makes the system easy to use and capable of generating photorealistic models with a high capacity to generate high quality textures.

In the future it will be necessary to evaluate the system in larger and more complex case studies and also indoors, where the results are not likely to be optimal, given the limitations of textureless photogrammetry. It will also be interesting to evaluate the system with higher resolution cameras.

REFERENCES

- Barua, Bhaskar, Clarence Gomes, Shubham Baghe, y Jignesh Sisodia. 2019. «A Self-Driving Car Implementation using Computer Vision for Detection and Navigation». Pp. 271-74 en 2019 International Conference on Intelligent Computing and Control Systems (ICCS). Madurai, India: IEEE.
- Cerrillo-Cuenca, Enrique, Pedro Ortiz-Coder, y José-Ángel Martínez-del-Pozo. 2014. «Computer Vision Methods and Rock Art: Towards a Digital Detection of Pigments». *Archaeological and Anthropological Sciences* 6(3):227-39. doi: 10.1007/s12520-013-0147-2.
- Dorian Gálvez-López, Marta salas, y Juan D tardós. s. f. «Real-time monocular object SLAM». *Robotics and Autonomous Systems*.
- James, Mike R., Stuart Robson, y Mark W. Smith. 2017. «3-D Uncertainty-Based Topographic Change Detection with Structure-from-Motion Photogrammetry: Precision Maps for Ground Control and Directly Georeferenced Surveys: 3-D Uncertainty-Based Change Detection for SfM Surveys». *Earth Surface Processes and Landforms* 42(12):1769-88. doi: 10.1002/esp.4125.
- Luhmann, Thomas, Stuart Robson, Stephen Kyle, y IA Harley. 2006. *Close Range Photogrammetry: Principles, Techniques and Applications*.
- Mikhail, E. M., J. S. Bethel, y J. C. McGlone. 2001. *Introduction to modern photogrammetry*. Wiley.
- Ortiz-Coder, Pedro, y Alonso Sánchez-Ríos. 2019. «A Self-Assembly Portable Mobile Mapping System for Archeological Reconstruction Based on VSLAM-Photogrammetric Algorithm». *Sensors* 19(18):3952. doi: 10.3390/s19183952.
- Ortiz-Coder, Pedro, y Alonso Sánchez-Ríos. 2020. «An Integrated Solution for 3D Heritage Modeling Based on Videogrammetry and V-SLAM Technology». *Remote Sensing* 12(9):1529. doi: 10.3390/rs12091529.
- Pollefeys, M., D. Nistér, J. M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S. J. Kim, P. Merrell, C. Salmi, S. Sinha, B. Talton, L. Wang, Q. Yang, H. Stewénus, R. Yang, G. Welch, y H. Towles. 2008. «Detailed Real-Time Urban 3D Reconstruction from Video». *International Journal of Computer Vision* 78(2-3):143-67. doi: 10.1007/s11263-007-0086-4.
- Raúl Mur-Artal, J. M. M. Montiel, y Juan D. Tardós. 2017. «ORB-SLAM: A Versatile and Accurate Monocular SLAM System». *IEEE Transactions on Robotics* 31(5):1255-62.
- Remondino, F., E. Nocerino, I. Toschi, y F. Menna. 2017. «A CRITICAL REVIEW OF AUTOMATED PHOTOGRAMMETRIC PROCESSING OF LARGE DATASETS». *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-2/W5:591-99*. doi: 10.5194/isprs-archives-XLII-2-W5-591-2017.
- Rupnik, Ewelina, Mehdi Daakir, y Marc Pierrot Deseilligny. 2017. «MicMac – a Free, Open-Source Solution for Photogrammetry». *Open Geospatial Data, Software and Standards* 2(1). doi: 10.1186/s40965-017-0027-2.
- Schöps, Thomas, Torsten Sattler, Christian Häne, y Marc Pollefeys. 2017. «Large-Scale Outdoor 3D Reconstruction on a Mobile Device». *Computer Vision and Image Understanding* 157:151-66. doi: 10.1016/j.cviu.2016.09.007.
- Torresani, A., y F. Remondino. 2019. «VIDEOGRAMMETRY VS PHOTOGRAMMETRY FOR HERITAGE 3D RECONSTRUCTION». *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-2/W15:1157-62*. doi: 10.5194/isprs-archives-XLII-2-W15-1157-2019.
- Zhang, Z. 2000. «A flexible new technique for camera calibration». *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(11):1330-34. doi: 10.1109/34.888718.
- Zlot, Robert, Michael Bosse, Kelly Greenop, Zbigniew Jarzab, Emily Jukes, y Jonathan Roberts. 2014. «Efficiently Capturing Large, Complex Cultural Heritage Sites with a Handheld Mobile 3D Laser Mapping System». *Journal of Cultural Heritage* 15(6):670-78. doi: 10.1016/j.culher.2013.11.009.