

ACCURATE 3D RECONSTRUCTION USING A VIDEOGRAMMETRIC DEVICE FOR HERITAGE SCENARIOS

P.Ortiz-Coder¹, Rubén Cabecera^{1,2},

¹ 2Freedom Imaging Hardware and Software S.L. – info@2freedom.net, rcs@aiuestudio.com

² University of Extremadura; Santa Ana University Centre, Department of Structures, Agricultural and Food Industries Engineering, Spain – rcs@unex.es

KEY WORDS: Videogrammetry, 3D Dense reconstruction, handheld scanner, automatic photogrammetry, Video 3d reconstruction, keyframe.

ABSTRACT:

In recent years, handheld laser scanning systems have been developed for documenting architectural heritage, among other applications. In this article we present a new alternative for the 3D documentation of historical heritage based on videogrammetry. For this purpose, a prototype has been designed with two cameras, a high resolution camera and a VGA camera which, when connected to a tablet, allow the user to establish a guidance system to ensure that the trajectory is not lost and enables highly flexible and long-lasting movements over time. This paper unveils the operation of the filtering and image selection system to finally proceed to an evaluation of the prototype in three areas of an archaeological site, called "Casa del Mitreo" in the city of Mérida (Spain). The results are compared to the Faro Focus 3D X330 laser scanner, yielding very similar accuracies and a capture time about 17 times faster than the 3D laser scanner. The article therefore proposes a real alternative to 3D data acquisition systems in applications for the graphic documentation of architectural and archaeological heritage.

1. INTRODUCTION

Terrestrial photogrammetry has evolved a lot in recent years, allowing more flexibility in data acquisition and better results. In recent years, successful homologous point finding has been achieved among images with very different perspectives (Cao et al., 2010; Lowe, 1999), even taken at different times and under different conditions, and these can then be used for orientation with progressively better accuracies. But there have also been significant improvements in 3D reconstruction, including textureless areas (Tan, 2016). Likewise, with self-calibration or improvements in bundle adjustment (Triggs et al., 2000), where very positive and exciting results can be observed for the present and future of this field of knowledge.

These scientific and technical improvements in photogrammetry have provided a further advance in the democratisation of the technology, which can be used by people without a particularly deep knowledge of the technique, using commercial sensors, even mobile phones, without the need for calibration or special care during data acquisition. It should be mentioned at this point that the improvement in the quality of images and the increased resolution of sensors (Gruen and Akca, 2008; Sirmacek and Lindenbergh, 2014), even in the above-mentioned mobile phones, have also facilitated this democratisation process.

In either case, in general terms, poor data acquisition affects the quality of the models (Luhmann et al., 2006); it can lead to deformations or the impossibility of orientation and reconstruction, if minimum standards in data acquisition are not observed. So far, in photogrammetry, image processing follows data acquisition, so that it is generally not possible to know whether the capture has been successful until it has been processed.

This paper focuses on ground type captures where the user carries the passive sensor while walking or driving a vehicle.

1.1 Videogrammetry

In order to facilitate data capture in photogrammetry, it has been demonstrated that this can be obtained through video. Videogrammetry for this type of capture requires, therefore, a process of extraction and selection of frames for the subsequent photogrammetric process. The methodology for selecting keyframes from among all the frames generated by the video can be divided into three types (Ortiz-Coder and Sánchez-Ríos, 2020; Torresani and Remondino, 2019):

- Constant selection. This methodology is the simplest and selects one out of n frames.
- 2D-Based Selection. This system calculates the homologous points between correlated images and selects the most suitable one through the number of tie points.
- 3D-Based Selection. It is the most complex and complete of all, since it performs a bundle adjustment of the frames in real time and selects the keyframes based on the geometry of the object. (Raúl Mur-Artal et al., 2017).

Video capture avoids human interaction and increases the possibilities of photogrammetry. But at the same time it has some challenges that we will mention below and that we have tried to solve in our proposal:

- The resolution of video cameras is generally lower than that of still cameras. And lower resolution affects the quality of the model. In addition to this problem, it is common in commercial video cameras for the sensors to be of the rolling shutter type, which can cause image deformation in mobile captures, i.e. while the camera is moving or while the user is walking.
- The geometry of the shot in video captures may not be as orderly as that taken manually, since it responds to

the continuous movement of the user as they move, often rather erratically, while walking, and conditionally while driving.

- If the video camera were to make excessively fast or inappropriate movements (e.g. self-rotations) in terms of the geometry of the shot, it would cause deformations or even the orientation of the images could not be solved later, generating different parts or, simply, loss of information. An added challenge is to gain reliability in this sense, being able to know in real time if this risk exists and how to solve it, without having to wait for post-processing, where the user may no longer be in the scene to be documented.

1.2 Previous work

In the scientific literature we can find different works with the objective of generating 3D models automatically through videos (Gruen, 1997; Liu et al., 2015; Torresani and Remondino, 2019). In the pursuit of this objective, we find two well differentiated lines of research:

A first line prioritises the generation of 3D models with a minimum computational cost and in the shortest possible time. In this work, accuracy is not a priority, using low resolution cameras and 3D reconstruction algorithms in real time or close to real time (Pollefeys et al., 2008). In this line of development we find the V-SLAM algorithms (Taketomi et al., 2017; Yousif et al., 2015) which were developed by the computer vision community for real-time robot localisation, but which have made the leap to be used in multiple applications, for example as an assistant in autonomous cars (Barua et al., 2019) or in laser-based 3D data acquisition systems (Leica Geosystem, 2021).

The second line of research uses the video frames to make more robust bundle adjustments, using the maximum resolution of the images, with the purpose of achieving maximum precision, even if the processing time increases. In this line are 3D measurement systems in engineering, architecture or heritage, among others (Ortiz-Coder and Sánchez-Rios, 2019; Torresani et al., 2021). In this field it is important to have high accuracy, over and above the post-processing time. Commercial mobile mapping systems, mainly based on lasers such as Zeb Horizon, manufactured by Geoslam Limited (Nottingham, England), BLK2GO manufactured by Leyca Geosystem (Heerbrugg, Switzerland), Vexcel Ultracam Pather manufactured by Vexcel Imaging (Anzengruebergasse, Austria), or the Stencil 2 manufactured by Kaarta, Inc (Pittsburgh, United State), use videogrammetric systems integrated with lasers for this purpose, where videogrammetry has a very variable real weight in the achievement of final results.

In terms of image-based handheld 3D model generation systems, there are some solutions on the market such as the ZED System from the company Stereolabs, Inc (San Francisco, United States), based on a stereocamera that can calculate the position of the camera in real time, as well as generate 3D models or apply object tracking or artificial intelligence systems. On the other hand, we find interesting works of data capture through video captured with mobile systems and using cloud processing for the generation of an initial sparse point cloud and a subsequent dense 3D point cloud (Nocerino et al., 2017). Another handheld system for image-based reconstruction with a certain similarity to the one proposed in this paper is the prototype presented by Torresani

(Torresani et al., 2021) which consists of two global shutter cameras with a certain convergence of the optical axes of the cameras, and which uses stereoscopy to perform image corrections. The guidance system is based on V-SLAM to finally generate the 3D models using photogrammetry.

1.3 Our proposal

In the scientific writing, as we have seen, it is common to find cameras with low resolution or inadequate sensors for subsequent orientation and dense point cloud generation in a post-processing phase (e.g. rolling shutter sensors).

This prototype proposal was made with the aim of achieving a dynamic data acquisition system that the user could use with maximum flexibility of movement and a guarantee that the data captured in the field would generate an accurate 3D model without causing breaks or disconnections in the trajectory. For this purpose, the system has a real-time V-SLAM algorithm which, in addition to different guidance tools, also in real time, helps the user to achieve this objective. For a better visualisation of the data and freedom of movement of the camera and the user, an external tablet is used to process and visualise the information in real time.

In our proposal we have opted to maximise the quality and resolution of the cameras and lenses, avoiding blurred or deformed images. We have also chosen a monocular solution, with the aim of achieving maximum precision, or, in other words, not losing this precision in an attempt to apply a scale to the model which, depending on the technique used, could compromise the absolute precision required for the average distance of the applications for which it is designed. The application of this system establishes a very variable range of distances, from 1m to 30m, among other applications such as engineering, structures, archaeology or constructions, etc.

Our proposal, therefore, is based on two cameras, one at higher resolution for the generation of 3D models through photogrammetry, and another camera at lower resolution to apply a real-time guidance system. The camera is carried by the user who moves freely around the scene. A tablet, connected by USB3 to the cameras, through software programmed for this purpose, saves the high resolution images and processes the signal from the lower resolution camera for the guidance system and assists the user while performing the data acquisition process. Subsequently, the data is processed through a specific keyframe selection and filtering system, so as to ensure the connection between all images and thus a continuous image orientation. Finally the high density 3d point cloud is generated using various photogrammetric algorithms.

Paper contributions

We can summarise the contribution of this paper in the following points:

- A reliable data acquisition system is presented, so that the capture in the field guarantees good subsequent results. This is achieved thanks to the interaction of different systems: V-SLAM-based guidance system, and keyframe filtering and selection system.
- The combination of the above factors allows the use of these systems by non-specialists, democratising the technology.
- The choice of the monocular option and separation from the computer system facilitates flexibility in use,

allowing it to be used on poles, or in areas that are difficult to access.

- A very flexible acquisition system is presented which tries to increase the capabilities of traditional photogrammetry by means of long period captures and high accuracies.

The previous publications (Ortiz-Coder and Sánchez-Ríos, 2019; Ortiz-Coder and Sánchez-Ríos, 2020) together with recent work carried out with the system proposed in this paper demonstrate that image-based 3D acquisition systems have the capability to generate highly accurate models in handheld type captures. Also, it is necessary to mention the known limitations of photogrammetry with respect to homogeneous (Ley et al., 2016) or reflective areas (Luhmann et al., 2006).

2. PROTOTYPE PROPOSAL

2.1 Hardware configuration

The designed prototype consists of two cameras: Camera A will be used for the real-time guidance system, so the camera chosen is ELP-USB500W05G, manufactured by Ailipu Technology Co., Ltd. (Shenzhen, Guangdong, China). This camera has a VGA resolution of 640x480 pixels, 25 fps framerate, and a field of view of 170°. For the photogrammetric acquisition, we have selected camera B which brand and model is JAI GO-5000-C from JAI Ltd. (Copenhagen, Denmark). This camera has a resolution of 2560 × 2048, a 1" sensor, a used framerate of 6fps and a field of view of 89° × 76.2°.

Both cameras have been calibrated internally (Zhang, 2000) and an external camera calibration has also been performed, calculating the position of one camera in relation to the other. The external calibration was performed by taking different shots of a corner with 15 targets measured with a total station with angular accuracy of 1" and 1.5 mm + 2 ppm distance error. A seven-parameter 3D transformation (Mikhail et al., 2001) was then performed to calculate the relative positions of each camera.

Both cameras are encapsulated in a casing, arranged side by side with the optical axes parallel to each other and with a baseline of 6cm, although this value has varied slightly in the different prototypes. The body has been 3D designed and printed on a 3D printer using ABS plastic. The design has evolved over the last few months and it has now been possible to optimise the space and stylise the lines to form a current prototype.

Finally, the cameras are synchronised in image capture, timing each frame with millisecond precision for both cameras, and estimating the offset and applying the appropriate corrections.

2.2 V-SLAM Software user's Guidance

One of the objectives in the design of our proposal was to increase the reliability of the data capture to guarantee that in the post-processing there would be no deformations or breakpoints in the trajectory. To this end, we use a V-SLAM (Raúl Mur-Artal et al., 2017) algorithm that calculates the bundle adjustment in real time and estimates initial keyframes that will serve as a starting point for the final selection of keyframes to be made during the subsequent selection process. Since our system has a camera A for the guidance system and a camera B for the acquisition of high resolution images for the photogrammetric process, the keyframe selection process established by the V-

SLAM system must be checked in the selection process. But also due to the results of previous experiments carried out with our prototype, we have observed that the three-dimensional models generated after applying our image selection algorithm increase the accuracy with respect to those that have used only those suggested by the initial process; we have even observed that in complex paths, the initial selection of keyframes is frankly deficient, causing breaks in the trajectory.

The guidance system is completed visually with image tracking and 3D sparse point cloud generation, as well as certain warnings to the user to avoid very homogeneous textures or to help the user to find the tracking again under circumstances where it is missing.

2.3 KeyFrames selection Filter

Once the data acquisition process has been completed, keyframe selection is performed. The algorithm tries to minimise the number of disconnected images and thus minimise the number of unconnected blocks or, as we have called them, segments.

Initially, and using the fine synchronisation between the two cameras, we use the first keyframes calculation performed by the V-SLAM during the data acquisition process to identify the images captured at the same time by camera B. And we apply a specially designed algorithm to calculate the homologous points between consecutive images using the SURF (Rubblee et al., 2011) descriptor, whereby, we analyse whether the number of calculated homologous points is greater than an estimated tolerance value. If the value is greater, the analysed image is inserted in the previous segment, but if the number of homologous points is less than a tolerance value, we will insert a number n of intermediate high resolution images before recalculating the homologous points. If the resulting value is still less than a value to the tolerance value, a new segment will be formed starting with this disconnected image, otherwise the algorithm will continue.

This keyframe selection process can be graphically represented in figure 1, and results in one or more segments whose internal connection is guaranteed. In our experience, the combination of the capture guidance system and the filtering system greatly minimises the possibility of more than one segment appearing, increasing the possibilities of capturing complex areas and capturing over long periods of time.

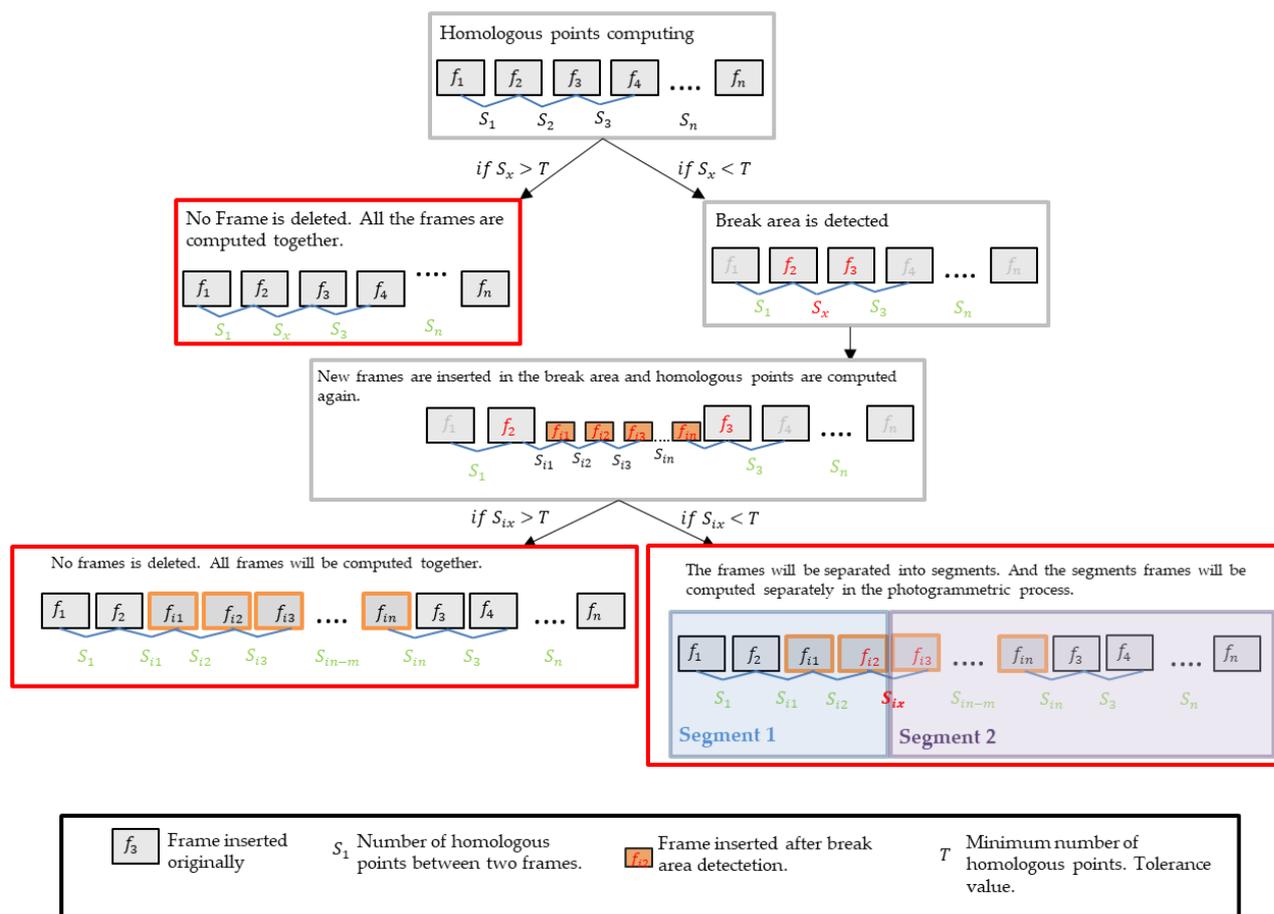


Figure 1. Key frames selection and filtering algorithm diagram.

3. EXPERIMENTS

2.4 3D color points cloud generation

Each calculated segment is oriented using a first direct calculation of the orientation parameters and, subsequently, using these parameters as initial values for an iterative bundle adjustment, using MICMAC (Rupnik et al., 2017). Subsequently, the point cloud is generated by calculating the depth maps for each main image, for which we have selected one image out of eight secondary images, four on each side of the main image (Pierrot Deseilligny and Clery, 2012). The point cloud has been filtered using a combination of filters called Radius Outlier Removal (ROR) and Statistical Outlier Removal (SOR) (Cignoni et al., 2008). The mesh of the point cloud was calculated using the Poisson algorithm (Hoppe, 2008) and for the texturing of the mesh a rectification of each main image (Ranzuglia et al., 2013) that would be used only to apply the texture was initially performed. The relative positions of each image were also considered to perform the projection of the images on the mesh.

Finally, and in order to join the different segments in case there is more than one within the same reference system, we use the trajectory calculated during the data capture process to perform a least squares adjustment (Mikhail et al., 2001) of the positions of each camera, calculated with the photogrammetric procedure, on those calculated in the V-SLAM process. Hence, it can be guaranteed that all segments will have a common reference system, even if the joint accuracy between segments is not of the same quality as that calculated with photogrammetric procedures.

In order to verify the proposed prototype and its capabilities in an architectural heritage environment, three areas of the so-called Casa del Mitreo, in the city of Mérida (Spain), have been documented. The Casa del Mitreo is a Roman house dating from between I BC and II AD. The house has several rooms of very different types, with large and well-preserved mosaics on the floors and colourful paintings on the walls. To test our prototype we have chosen three different scenarios:

- Area 1: Pond and Peristylum. This area has columns and zones at different levels, so we will test the flexibility in catching in diverse and complex circumstances.
- Area 2: Underground rooms. In this area there are two levels of height, communicated by a staircase and very narrow areas, with a great difference in light between the different areas. Therefore, we will be able to test the functioning of the prototype to adapt to the different light conditions and to adapt to very narrow spaces.
- Area 3: Rooms with mosaic floors. This area consists of several rooms, three of them with mosaics, mainly of a geometric nature. In addition to testing the accuracy of the system in larger areas, we will test the performance with repetitive geometric captures that could, a priori, compromise the correct orientation of the images.

In order to compare the generated data with our prototype, we used the FARO Focus3D X330 laser scanner to document the three scenarios as well. The internal camera of the scanner was

used to obtain the colour of the point cloud from each scanning position. With the same objective, measurements were taken with a Pentax V-227N (Pentax Ricoh Imaging Company, Ltd., Tokyo, Japan) prismless total station of 180 targets distributed equally between the three study areas.

The documentation of the three scenarios using the proposed prototype is done while walking through them, holding the system in one hand and the Tablet (HP Pavillion Intel Core i5-8250U 14", with a 256 GB SSD and 16 GB RAM, manufactured by HP Inc. (Palo Alto, CA, USA)) with the guidance system in the other (see figure 2). The guidance system indicates to the user if he/she has made an inappropriate movement and has lost the tracking, as well as when he/she recovers it, displaying everything on the screen.

The time used for data acquisition with the proposed prototype has fluctuated between 8 and 12 minutes, depending on the test area. The trajectory followed has tried to cover the whole scanning area, as well as to start and end in the same area, in order to minimise the error by applying the loop closure algorithm (Ortiz-Coder and Sánchez-Rios, 2019). The graph of the trajectories can be seen in figure 3. As can be seen, these follow a complex and variable distribution, totally adapted to the conditions of each scene, demonstrating the flexibility that this system can achieve.

Scanning time (minutes)/Area	Laser scanner Faro X330	Proposed system
Area 1	200'	11'
Area 2	100'	8'
Area 3	220'	12'

Table 1. Data acquisition time used by the proposed prototype and the laser scanner Faro Focus3d X330.

3.1 Results

Once all the data had been processed with both systems, the average resolution of the point clouds generated with both methods was calculated in the three study areas, including overlap areas, resulting in an average point-to-point distance of 0.4mm for the Faro Focus 3D X330 laser scanner and 1.4mm for the proposed prototype.

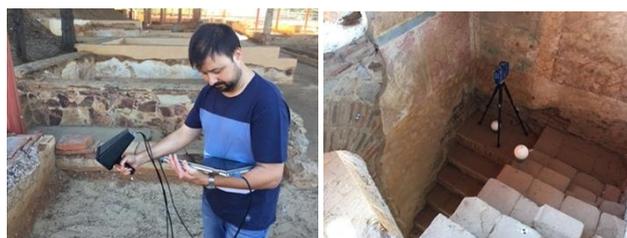


Figure 2. Data acquisition in the "Casa del Mitreo": Data acquisition using the proposed prototype (left); Laser scanning capture process using Faro Focus X330 (right).

Using the targets measured with the total station, Euclidean average distance error (δ_{avg}) and the Root Mean Square Error (RMSE) (Hong et al., 2015) have been calculated for the models generated with our system and with the models generated with the laser scanner. The results can be seen in table 2, where the RMSE has been calculated for the three components X, Y, X. As

can be seen, the errors are very similar between both capture systems.

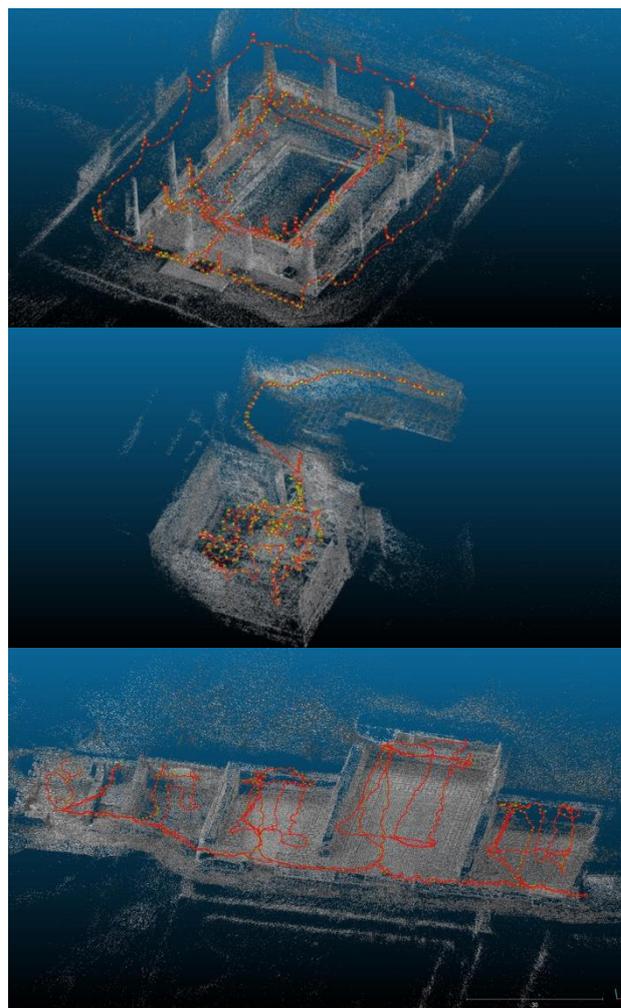


Figure 3. Trajectories followed to document the three study areas. Area 1 (up), area 2 (middle), area 3 (bottom). The red line defines the trajectory followed and the green dots, the positions of the Keyframes calculated.



Figure 4. 3D mesh textured of the Pond and Peristylum (area 1)(up) and of the underground room(area 2) (bottom).

REFERENCES

	Faro Focus3D X 330				Proposed System			
Area 1								
	Vector X	Vector Y	Vector Z	Error	Vector X	Vector Y	Vector Z	Error
δ_{avg}				10				7
RMSE	9	6	4	7	6	4	4	5
Area 2								
	Vector X	Vector Y	Vector Z	Error	Vector X	Vector Y	Vector Z	Error
δ_{avg}				8				8
RMSE	5	9	4	6	4	6	6	6
Area 3								
	Vector X	Vector Y	Vector Z	Error	Vector X	Vector Y	Vector Z	Error
δ_{avg}				11				8
RMSE	5	8	7	6	6	4	5	5

Table 2. Errors made by the laser scanner(left) and our proposed system (right) process. Euclidean average distance error(δ_{avg}) and Root Mean Square Error (RMSE) have been calculated for the three areas of study.

4. CONCLUSIONS

In this article we present a novel 3D scanning system based on videogrammetry. The instrument consists of three basic keys: I. A monocular system with high quality and high-resolution cameras, II. A real-time user guidance and assistance system, and III. An image selection and filtering system according to the hardware and capture method used.

Our proposal makes it possible to capture images over a long period of time without the danger of trajectory breakage. But they also ensure that the data is processed using photogrammetric formulations without generating new blocks or unconnected parts.

The nature of the proposal allows for a very flexible capture, enabling a wide variety of movements, as well as adapting to the scanned object by moving closer or further away, among other movements.

The speed of capture and the accuracies achieved by the proposed system in these case studies in historical heritage applications clearly open up an alternative to current scanning systems, such as the laser scanner, as has been demonstrated in this article.

For future work it would be necessary to implement a system that would provide the models with a metric in real dimensions, whether based on IMU's, GNSS-GPS or other instruments.

Barua, B., Gomes, C., Baghe, S., Sisodia, J., 2019. A Self-Driving Car Implementation using Computer Vision for Detection and Navigation, in: 2019 International Conference on Intelligent Computing and Control Systems (ICCS). Presented at the 2019 International Conference on Intelligent Computing and Control Systems (ICCS), IEEE, Madurai, India, pp. 271–274.
<https://doi.org/10.1109/ICCS45141.2019.9065627>

Cao, B., Ma, C., Liu, Z., 2010. Affine-invariant SIFT descriptor with global context, in: 2010 3rd International Congress on Image and Signal Processing. Presented at the 2010 3rd International Congress on Image and Signal Processing (CISP), IEEE, Yantai, China, pp. 68–71.
<https://doi.org/10.1109/CISP.2010.5647071>

Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., Ranzuglia, G., 2008. MeshLab: an Open-Source Mesh Processing Tool. Eurographics Ital. Chapter Conf. 8 pages.
<https://doi.org/10.2312/LOCALCHAPTEREVENTS/I TALCHAP/ITALIANCHAPCONF2008/129-136>

Gruen, A., 1997. Fundamentals of videogrammetry — A review. Hum. Mov. Sci. 16, 155–187.
[https://doi.org/10.1016/S0167-9457\(96\)00048-6](https://doi.org/10.1016/S0167-9457(96)00048-6)

Gruen, A., Akca, D., 2008. Evaluation of the metric performance of mobile phone cameras. Conference: 6th International ARIDA (Association for Real Time Imaging and Dynamic Analysis) Workshop "Innovation in 3D measurement, modeling and visualization" At: Trento, Italy

Hong, S., Jung, J., Kim, S., Cho, H., Lee, J., Heo, J., 2015. Semi-automated approach to indoor mapping for 3D as-built building information modeling. Comput. Environ. Urban Syst. 51, 34–46.
<https://doi.org/10.1016/j.compenurbysys.2015.01.005>

Hoppe, H., 2008. Poisson surface reconstruction and its applications, in: Proceedings of the 2008 ACM Symposium on Solid and Physical Modeling - SPM '08. Presented at the the 2008 ACM symposium, ACM Press, Stony Brook, New York, p. 10.
<https://doi.org/10.1145/1364901.1364904>

Ley, A., Hänsch, R., Hellwich, O., 2016. RECONSTRUCTING WHITE WALLS: MULTI-VIEW, MULTI-SHOT 3D RECONSTRUCTION OF TEXTURELESS SURFACES. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. III-3, 91–98. <https://doi.org/10.5194/isprs-annals-III-3-91-2016>

Liu, X., Tong, X., Yin, X., Gu, X., Ye, Z., 2015. Videogrammetric technique for three-dimensional structural progressive collapse measurement. Measurement 63, 87–99.
<https://doi.org/10.1016/j.measurement.2014.11.023>

Lowe, D.G., 1999. Object recognition from local scale-invariant features, in: Proceedings of the Seventh IEEE International Conference on Computer Vision.

- Presented at the Proceedings of the Seventh IEEE International Conference on Computer Vision, IEEE, Kerkyra, Greece, pp. 1150–1157 vol.2. <https://doi.org/10.1109/ICCV.1999.790410>
- Luhmann, T., Robson, S., Kyle, S., Harley, I., 2006. Close Range Photogrammetry: Principles, Techniques and Applications.
- Mikhail, E.M., Bethel, J.S., McGlone, J.C., 2001. Introduction to modern photogrammetry. Wiley
- Nocerino, E., Poiesi, F., Locher, A., Tefera, Y.T., Remondino, F., Chippendale, P., Van Gool, L., 2017. 3D RECONSTRUCTION WITH A COLLABORATIVE APPROACHBASED ON SMARTPHONES AND A CLOUD-BASED SERVER. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* XLII-2/W8, 187–194. <https://doi.org/10.5194/isprs-archives-XLII-2-W8-187-2017>
- Ortiz-Coder, P., Sánchez-Ríos, A., 2020. An Integrated Solution for 3D Heritage Modeling Based on Videogrammetry and V-SLAM Technology. *Remote Sens.* 12, 1529. <https://doi.org/10.3390/rs12091529>
- Ortiz-Coder, P., Sánchez-Ríos, A., 2019. A Self-Assembly Portable Mobile Mapping System for Archeological Reconstruction Based on VSLAM-Photogrammetric Algorithm. *Sensors* 19, 3952. <https://doi.org/10.3390/s19183952>
- Pierrot Deseilligny, M., Clery, I., 2012. APERO, AN OPEN SOURCE BUNDLE ADJUSTMENT SOFTWARE FOR AUTOMATIC CALIBRATION AND ORIENTATION OF SET OF IMAGES. *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* XXXVIII-5/W16, 269–276. <https://doi.org/10.5194/isprsarchives-XXXVIII-5-W16-269-2011>
- Pollefeys, M., Nistér, D., Frahm, J.-M., Akbarzadeh, A., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Kim, S.-J., Merrell, P., Salmi, C., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewénius, H., Yang, R., Welch, G., Towles, H., 2008. Detailed Real-Time Urban 3D Reconstruction from Video. *Int. J. Comput. Vis.* 78, 143–167. <https://doi.org/10.1007/s11263-007-0086-4>
- Ranzuglia, G., Callieri, M., Dellepiane, M., Cignoni, P., Scopigno, R., 2013. MeshLab as a complete tool for the integration of photos and color with high resolution 3D geometry data. *CAA 2012 Conf. Proc.*, Pallas Publications - Amsterdam University Press (AUP) 406–416.
- Raúl Mur-Artal, J. M. M. Montiel, Juan D. Tardós, 2017. ORB-SLAM: A Versatile and Accurate Monocular SLAM System. *IEEE Trans. Robot.* 31, 1255–1262.
- Rublee, E., Rabaud, V., Konolige, K., Bradski, G., 2011. ORB: An efficient alternative to SIFT or SURF, in: 2011 International Conference on Computer Vision. Presented at the 2011 IEEE International Conference on Computer Vision (ICCV), IEEE, Barcelona, Spain, pp. 2564–2571. <https://doi.org/10.1109/ICCV.2011.6126544>
- Rupnik, E., Daakir, M., Pierrot Deseilligny, M., 2017. MicMac – a free, open-source solution for photogrammetry. *Open Geospatial Data Softw. Stand.* 2. <https://doi.org/10.1186/s40965-017-0027-2>
- Sirmacek, B., Lindenbergh, R., 2014. Accuracy assessment of building point clouds automatically generated from iphone images. *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* XL-5, 547–552. <https://doi.org/10.5194/isprsarchives-XL-5-547-2014>
- Taketomi, T., Uchiyama, H., Ikeda, S., 2017. Visual SLAM algorithms: a survey from 2010 to 2016. *IPSIJ Trans. Comput. Vis. Appl.* 9, 16. <https://doi.org/10.1186/s41074-017-0027-2>
- Tan, D., 2016. Image Enhancement Based on Adaptive Median Filter and Wallis Filter, in: Proceedings of the 2015 4th National Conference on Electrical, Electronics and Computer Engineering. Presented at the 2015 4th National Conference on Electrical, Electronics and Computer Engineering, Atlantis Press, Xi'an, China. <https://doi.org/10.2991/nceee-15.2016.142>
- Torresani, A., Menna, F., Battisti, R., Remondino, F., 2021. A V-SLAM Guided and Portable System for Photogrammetric Applications. *Remote Sens.* 13, 2351. <https://doi.org/10.3390/rs13122351>
- Torresani, A., Remondino, F., 2019. VIDEOGRAMMETRY VS PHOTOGRAMMETRY FOR HERITAGE 3D RECONSTRUCTION. *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* XLII-2/W15, 1157–1162. <https://doi.org/10.5194/isprs-archives-XLII-2-W15-1157-2019>
- Triggs, B., McLauchlan, P., Hartley, R., Fitzgibbon, A., 2000. Bundle adjustment - A modern synthesis. *ICCV 99 Proc. Int. Workshop Vis. Algorithms Theory Pract.* 198–372.
- Yousif, K., Bab-Hadiashar, A., Hoseinnezhad, R., 2015. An Overview to Visual Odometry and Visual SLAM: Applications to Mobile Robotics. *Intell. Ind. Syst.* 1, 289–311. <https://doi.org/10.1007/s40903-015-0032-7>
- Zhang, Z., 2000. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* 22, 1330–1334. <https://doi.org/10.1109/34.888718>

APPENDIX

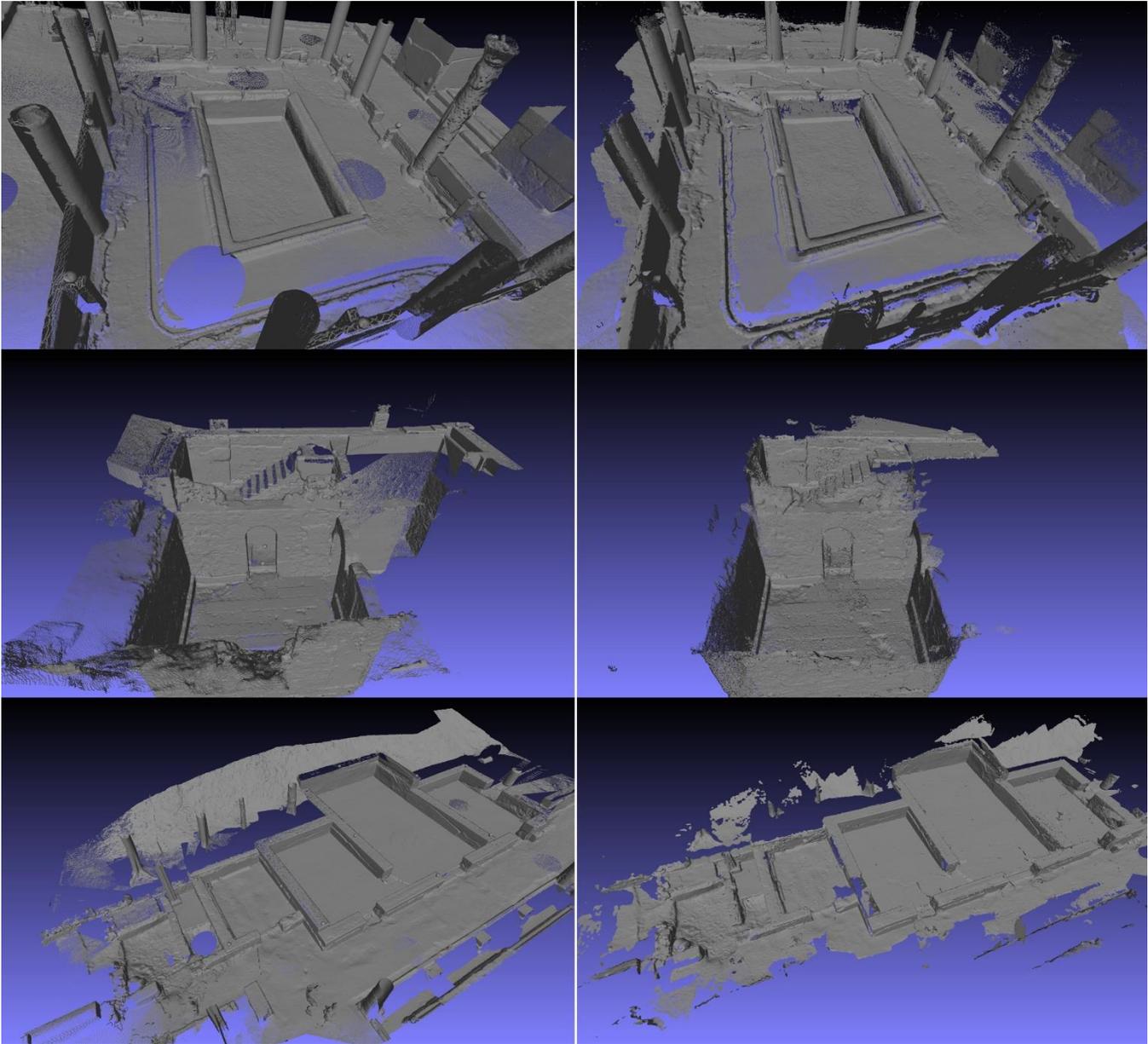


Figure 5. Visual comparison between 3D models without texture generated with the Faro Focus 3D X330 laser scanner (Left) and with the proposed system (Right) for areas 1(up), area 2(middle) and area 3 (bottom).