EXPERIMENTS USING SMARTPHONE-BASED VIDEOGRAMMETRY FOR LOW-COST CULTURAL HERITAGE DOCUMENTATION

A. Murtiyoso* and P. Grussenmeyer

Université de Strasbourg, INSA Strasbourg, CNRS, ICube Laboratory UMR 7357, Photogrammetry and Geomatics Group, 67000 Strasbourg, France - (arnadi.murtiyoso, pierre.grussenmeyer)@insa-strasbourg.fr

KEY WORDS: videogrammetry, photogrammetry, smartphone, low cost, heritage documentation

ABSTRACT:

The rapid development of 3D scanning technology is a welcome progress in the field of tangible cultural heritage documentation. While active sensors such as handheld Time-of-Flight (ToF) cameras and lidar have recently generated much hype, developments in low-cost imaging sensors have also seen long strides in recent decades. This paper aims to see the potential of videogrammetry for the purposes of heritage documentation. This technique has existed for decades, but we argue that when combined with modern smartphone sensors and proper photogrammetric processing workflow it may present an interesting low-cost solution for 3D scanning. Furthermore, the paper wishes to address the requirement for a certain geometric quality in heritage documentation and how the proposed method may fulfil them. For this reason, comparisons between the videogrammetric result and traditional DSLR close range photogrammetry will be described to determine its suitability for heritage documentation. Results show that using modern low-cost smartphone imaging sensors, a good compromise between geometric quality and overall cost in the context of cultural heritage recording is possible to achieve.

1. INTRODUCTION

Videogrammetry or video-based photogrammetry has existed for a long time. The idea of extracting video frames to be used as input to photogrammetry was developed together with the advent of low-cost CCD sensors (Gruen, 1997). Although the technique benefits greatly from developments in digital video sensors and image orientation algorithms (e.g. SLAM), it remains in the shadows of traditional image-based photogrammetry due to the inherent loss of resolution when passing from image to video.

The technique presents obvious advantages *vis-à-vis* classical photogrammetry. Acquisition time is greatly reduced because the user only needs to take a video shot compared to the more time-consuming multi-image acquisition. However, the quality of the resulting 3D data is often severely limited by the quality of the video and the frame sequencing itself (Flies et al., 2019).

As such, videogrammetry applications are usually limited to projects requiring less precision but higher difficulty for image acquisition e.g. fast moving objects (Bailey et al., 2020), underwater settings (Price et al., 2019), or complex medical operations (Lerma et al., 2018). In the field of heritage documentation, videogrammetry is an interesting method due to its low-cost nature (Ahmad et al., 2019; Sun and Zhang, 2019).

In this paper, we argue that state-of-the-art smartphone videos may be used as a viable alternative for heritage documentation, specifically for applications not requiring a high level of precision e.g. visualisation, VR, or AR. In light of the recent developments in other similarly low-to-medium cost and near-real time sensors (Wang et al., 2020), it is interesting to see how smartphone videogrammetry fares for the specific context of heritage documentation. A similar study was conducted by Torresani and Remondino (2019) in which the authors concluded that it is theoretically viable to use smartphone videogrammetry to generate 3D records for heritage objects, although they maintained that traditional photogrammetry is still irreplaceable for higher-precision requirements.

2. METHODOLOGY

In the experiments, we specifically tested the use of lower end smartphones to acquire videos to be used for the 3D reconstruction. These video-based reconstructions were compared to DSLR-based reconstructions and in some cases laser scans to determine its quality. The aim of the study is not to determine the most precise method as this is easily concluded just from the quality of the raster inputs and has been demonstrated by Flies et al. (2019). The main idea is therefore to verify if the acquired quality in terms of: (1) geometric precision, (2) textural quality, and (3) details represented is sufficient for specific heritage applications. The applications concern mainly those related to knowledge diffusion such as 3D visualisation, AR, and to some extent VR.

Aside from a general overview of videogrammetric results, the paper will also present the quantitative analysis of two specific case studies. Two types of heritage objects are discussed in this experiment: (1) an interior case study in the form of a marble statue and (2) an exterior case study in the form of church tympanums.

In both qualitative case studies, a video was taken using a Motorola G7 smartphone camera. First released in February 2019 with a retail price of 250 EUR, this series of Android smartphone consists of several individual types. For the purposes of the tests, a regular Moto G7 was used. The video quality used in the experiments is of a FHD (1080p) quality. Acquisition strategy amounted to the most important part of the videogrammetric workflow. In accordance to basic

* Corresponding author

photogrammetric principles, the video was taken in such way that it goes around the object in a convergent manner (see illustration in Figure 1). Shooting speed was another important factor during data acquisition. Slower video will enable more overlap between the eventual sequenced frames, but would naturally increase the amount of data to be processed. However, the sequencing rate is modifiable and therefore a slow video can eventually be divided into fewer frames when data size is a problem.

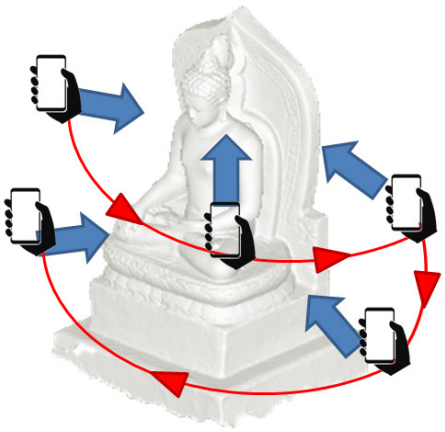

Figure 1. Illustration of the video acquisition method used in this paper. The main idea is to shoot a video converging around the object in question.

The videos were then processed using the commercial software Agisoft Metashape, by benefitting from its video sequencing functionality. Also in both quantitative case studies DSLR photogrammetry and terrestrial laser scanning (TLS) point clouds are available for comparison.

3. RESULTS AND DISCUSSIONS

3.1 Visual assessment

As has been previously mentioned, videogrammetry presents a distinct advantage in the form of very fast acquisition time. However, this must of course follow a strict guideline to ensure favourable geometric networks during the eventual image orientation process. This includes, among others, sufficient overlap between the images and a convergent image network. For the first element, the choice of frame extraction from the video is of the upmost importance. Sequences that are too far apart risk to reduce the overlap, however sequences that are too close together would increase processing time. As has been mentioned by other studies, the main obstacle to high-precision 3D reconstruction using videos is the quality of the frame images. Indeed, even in a high definition setting the possibility of having blurred frames is still present.

In Figure 2, several examples of 3D products are showcased both in the form of 3D point clouds and 3D mesh. All of these 3D models were generated from short videos ranging from 1 minute to 1 minute 10 seconds. They were shot by closely following photogrammetric requirements, i.e. covering the most of the object by trying to follow a convergent geometry.

(a) *(b)* *(c)* *(d)*

(e) *(f)*

Figure 2. A selection of examples of 3D models generated by the smartphone videogrammetry technique. First row shows a trio of objects displayed in the Lëtzebuerg City Museum, Luxembourg: (a) Gothic relief, (b) medieval tombstone, (c) wooden statue of a bishop. (d) shows the marble statue of “Rosa Mystica” in the church of St-Pierre-le-Jeune in Strasbourg, France. Second row shows exterior cases: (e) tympanum of the church of St-Jean in Strasbourg France, and (f) tympanum of the church of St-Pierre-le-Jeune.

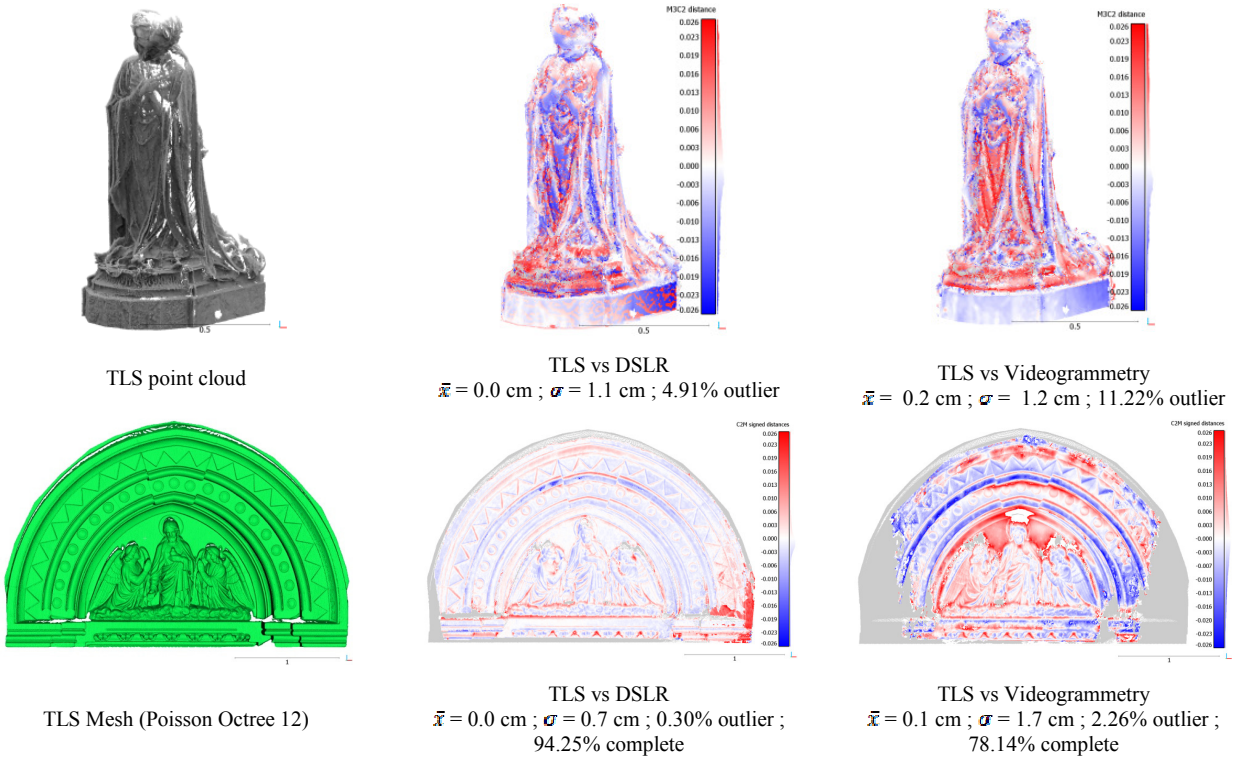

Figure 3. Geometric analysis for the data from Figure 2 (d) and (f). $\bar{x}$ denotes average deviation from TLS and σ the standard deviation. The outlier points are points with a deviation greater than the set tolerance. The completeness level is how close each respective point cloud covers the ideal closed surface as represented by the TLS mesh.

Dataset	Type	Video Duration	Extracted Frames
"Relief"	Interior	00:32 (70 MB)	38
"Tombstone"		00:35 (74 MB)	20
"Bishop"		01:01 (126 MB)	58
"Rosa"		00:26 (56 MB)	48
"St-Jean"	Exterior	00:52 (18 MB)	79
"St-Pierre"		00:21 (45 MB)	33

Table 1. Characteristics of the input video for the objects described in Figure 2. All videos were shot using a Moto G7 with an FHD 1080p resolution.

The first three objects in Figure 2 are collections of the Lëtzebuerg City Museum, Luxembourg while the fourth one is a statue inside the St-Pierre-le-Jeune church in Strasbourg, France. Together, the four objects in the first row show the application of this technique in indoor situations. The second row on the other hand, showcases results for an outdoor situation. This is particularly represented by tympanums of two Strasbourg churches: St-Jean and St-Pierre-le-Jeune. Furthermore, Table 1 describes the characteristics of the input video for each object displayed in Figure 2 including video duration and the number of frames thereafter extracted for the 3D reconstruction process. Note that all the videos were shot using the same smartphone (Moto G7) and the same parameters, with camera path more or less following those illustrated by Figure 1.

Results from Figure 2 show that the visual quality of the models is satisfactory and in most cases enough for visualisation purposes. Even in indoor settings, point cloud noise is minimal. Problems with the point cloud mostly came from the lack of data, as can be seen in Figure 2 (f). In this particular case, in addition to the relatively short duration of the input video (see Table 1), the sequenced frames presented many blurred photos which contributed to this result. The mesh models also presented sufficient quality, while the lack of some details as expected from the lower quality of the input data.

In order to perform a quantitative analysis, Figure 2 (d) and (f) were then compared to the available DSLR photogrammetry and TLS point clouds. The results of this comparison shall be described in the next section.

3.2 Qualitative assessment

In Figure 3, the 3D reconstruction from videogrammetry was compared against a TLS point cloud for the interior dataset and a TLS mesh for the outdoor dataset. The TLS used in this case is a FARO Focus X330. While higher quality laser scanner may be used in this case, for the objective overall detail resolution of 1 cm this type of scanner was deemed sufficient.

For comparison, point clouds of the same object generated by Canon EOS 6D images were also added to the figure to represent the more traditional method for 3D reconstruction of these types of cultural heritage objects. Using a pre-determined detail resolution (σ) of 1 cm, a tolerance for the point cloud deviation was set at 2.6 cm (2.6 σ hence 99% confidence level). Each point cloud was then registered to the TLS reference using the Iterative Closest Point (ICP) method.

Deviation analysis was performed using the open source software CloudCompare (https://www.danielgm.net/cc/, accessed 14 June 2021). The cloud-to-cloud Euclidean distance was computed for the interior dataset, while a cloud-to-mesh analysis was possible for the exterior dataset due to the presence of reference mesh data. In this regard, points within each point cloud that registers a Euclidean distance of more than the set

tolerance of 2.6 cm, regardless of the type of reference (point cloud of mesh) is considered as noise.

As can be seen on Figure 3, the metrics show that geometrically speaking this technique managed to deliver sufficiently accurate results with minor standard deviations when using TLS data as reference. In the interior "Rosa" dataset, the average error is only 2 mm compared to DSLR's 0 mm, while the difference in standard deviation between the two methods is negligible. Note that this statistical variable represents how much the non-noise points correspond to the reference data. Again, the problem as can be noticed is the level of detail on the objects which is not as high as the results from DSLR photogrammetry when compared to TLS data. The lower level of completeness in videogrammetric results can also be attributed to two factors: lack of data and noise filtering.

In the case of interior objects, the main concerns in photogrammetry were exacerbated with the use of videos. For example, the material of the object is often an important factor in performing 3D reconstruction using photogrammetry. The "Rosa" statue used in the interior case study is made of marble, a material which tends to be smooth. This lack of texture is always a challenge for dense matching algorithms (Murtiyoso et al., 2016).

In Figure 3, this is true for the DSLR result as much noise can already be observed visually. Quantitatively, this is represented by the outlier rate of points within the point cloud which amounted to almost 5%. This rate is quite high when compared to results from other types of materials such as stone. This is the case for the exterior data set ("St-Pierre") of which only 0.3% of points were considered as outliers. However, the level of noise in the videogrammetry result, while expected, was much higher with a rate of 11.22%. In other words, 1 in 10 points generated by the videogrammetric workflow was considered as noise.

For experiments regarding the exterior tympanum of "St-Pierre", again the non-noise points generated very good results with an average error of 1 mm for videogrammetry with slightly higher standard deviation (1.7 cm). Note that the tympanum is made from red sandstone. As has been previously hypothesised; the level of noise for this particular test is relatively low for both methods. The DSLR method only generated 0.3% points categorised as outlier, while using videogrammetry this value is slightly higher at 2.26%. Furthermore, in this experiment a mesh model was used as a reference which enabled the computation of the completeness of the 3D reconstruction *vis-à-vis* an ideal surface here represented by the TLS mesh.

The DSLR point cloud presented a very adequate completeness level, with 97.25% coverage of the ideal 3D mesh. The videogrammetry method, however, suffered much in this regard by registering only 78.14% with visually evident lack of points as shown in grey in Figure 3. As has been previously mentioned, this phenomenon is most likely caused by the lack of data. This lack of data due to two main factors: firstly, the low quality of the sequenced frames means that a large part of the tympanum was not reconstructed. This blurriness is due to the higher speed of video acquisition, hence reiterating the importance to strike a balance between acquisition speed and data storage size as mentioned previously in Section 2. Secondly, the duration of the input video itself as seen in Table 1 is relatively short. Indeed, in average 1.5 frames were extracted for each second of the video which may also contribute to the blurriness of the frames. While evidently this means more overlap is achieved for the dense matching, this further shows that over-sequencing of the video does not necessarily increase the final quality of the 3D model.

4. CONCLUSIONS AND PERSPECTIVES

This paper has attempted to describe the use of smartphone-based videogrammetry as a low cost solution to 3D scanning. As far as the visual aspect is concerned, Figure 2 showed that the technique managed to generate relatively good results. Nevertheless, the level of detail and noise level is still a problem although these are to be expected when considering basic photogrammetric and dense matching principles. That being said, the aim of the paper is to determine if such level of detail is enough for specific heritage documentation purposes.

More quantitative experiments also show that the results can achieve high accuracy when compared to a reference ground truth, in this case TLS-derived point cloud and mesh model. This is achieved in both our exterior and interior case studies, albeit with slightly differing standard deviation values. However, it is also worth of note that the value of average error does not reflect only the quality of the sensor but also that of the registration and/or georeferencing process. Indeed, by performing ICP for the registration process, the low average error is to be expected. Other statistics were therefore computed to better reflect the quality of the sensor, namely the outlier rate and the completeness rate. Results show that in interior setting with the marble Rosa statue, videogrammetry fared worse by generating more noise. This is somewhat remedied when facing richly-textured stone material as is the case of the exterior tympanum case study. However, another issue encountered in this case is the quality of the sequenced frames. A balance between overlap and frame quality is an important aspect in this regard.

Despite the shortcomings, we argue that good results are possible to be attained even by non-expert users. The caveat to this is of course the necessity to understand proper basic data acquisition strategies (i.e. converging video, slow enough shooting speed, lighting, etc.). A proper understanding of the objects properties (situation in exterior/interior, materials, etc.) also help in creating a good 3D model via videogrammetry.

Based on these results, we argue that the smartphone-based videogrammetry method has matured in the last decade. The geometric precision produced by this method is sufficient and the level of noise satisfactory even when performed using low-cost smartphones. Indeed, the main problem as can be seen from the results is the level of detail. We argue that for some visual-based applications of heritage documentation, this low-cost technique is today a viable alternative to other more expensive methods. Striking a balance between quality and cost in a low-cost range of prices may be a much easier consideration in any budget-tight projects as is often the case in heritage documentation. However, when considering high precision projects such as mapping, orthophoto, or CAD-based applications, image-based photogrammetry and by extension laser scanning produces undeniably better results. Users should therefore understand which techniques to be used for which purpose of heritage documentation.

ACKNOWLEDGEMENTS

The authors wish to thank the following people for their help in this research: Fr. Jérôme Hess and Marie Conrath of the St-Pierre-le-Jeune Catholic Parish, Strasbourg, France; the Lëtzebuerg City Museum, Luxembourg; and Dr. Anita Saraswati of the University of Luxembourg.

REFERENCES

Ahmad, N., Azri, S., Ujang, U., Cuétara, M.G., Retortillo, G.M., Mohd Salleh, S., 2019. Comparative analysis of various camera input for videogrammetry. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives 42, 63–70.

Bailey, A.M., Sherwood, C.P., Funk, J.R., Crandall, J.R., Carter, N., Hessel, D., Beier, S., Neale, W., 2020. Characterization of Concussive Events in Professional American Football Using Videogrammetry. Annals of Biomedical Engineering 48, 2678–2690.

Flies, M.J., Larsen, P.K., Lynnerup, N., Villa, C., 2019. Forensic 3D documentation of skin injuries using photogrammetry: photographs vs video and manual vs automatic measurements. International Journal of Legal Medicine 133, 963–971.

Gruen, A., 1997. Fundamentals of videogrammetry - A review. Human Movement Science 16, 155–187.

Lerma, J.L., Barbero-García, I., Marqués-Mateu, Á., Miranda, P., 2018. Smartphone-based video for 3D modelling: Application to infant’s cranial deformation analysis. Measurement: Journal of the International Measurement Confederation 116, 299–306.

Murtiyoso, A., Grussenmeyer, P., Koehl, M., Freville, T., 2016. Acquisition and Processing Experiences of Close Range UAV Images for the 3D Modeling of Heritage Buildings, in: Ioannides, M., Fink, E., Moropoulou, A., Hagedorn-Saupe, M., Fresa, A., Liestøl, G., Rajcic, V., Grussenmeyer, Pierre (Eds.), Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection: 6th International Conference, EuroMed 2016, Nicosia, Cyprus, October 31 -- November 5, 2016, Proceedings, Part I. Springer International Publishing, pp. 420–431.

Price, D.M., Robert, K., Callaway, A., Lo lacono, C., Hall, R.A., Huvenne, V.A.I., 2019. Using 3D photogrammetry from ROV video to quantify cold-water coral reef structural complexity and investigate its influence on biodiversity and community assemblage. Coral Reefs 38, 1007–1021.

Sun, Z., Zhang, Y., 2019. Accuracy evaluation of videogrammetry using a low-cost spherical camera for narrow architectural heritage: An observational study with variable baselines and blur filters. Sensors (Switzerland) 19.

Torresani, A., Remondino, F., 2019. Videogrammetry vs photogrammetry for heritage 3D reconstruction. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives 42, 1157–1162.

Wang, D., Watkins, C., Xie, H., 2020. MEMS mirrors for LiDAR: A review. Micromachines 11, 456.