

## A VERSATILE MULTI-CAMERA SYSTEM FOR 3D ACQUISITION AND MODELING

O. Lanz<sup>\* 1,6</sup>, F. Sottas<sup>2</sup>, M. Conni<sup>3</sup>, M. Boschetti<sup>1,4</sup>, E. Nocerino<sup>5</sup>, F. Menna<sup>6</sup>, F. Remondino<sup>6</sup>

<sup>1</sup> Covision Lab, Bressanone-Brixen, Italy - Email: <oswald.lanz><marco.boschetti>@covisionlab.com, Web: <https://covisionlab.com>

<sup>2</sup> Durst Phototechnik AG, Bressanone-Brixen, Italy

<sup>3</sup> Barbieri Electronics, Bressanone-Brixen, Italy

<sup>4</sup> Microtec, Bressanone-Brixen, Italy

<sup>5</sup> LIS UMR 7020, Aix-Marseille Université, CNRS, ENSAM, Université De Toulon, Marseille, 13397, France –  
erica.nocerino@univ-amu.fr

<sup>6</sup> Bruno Kessler Foundation (FBK), Trento, Italy – Email: <lanz><fmenna><remondino>@fbk.eu

### Commission II – WG II/8

**KEY WORDS:** photogrammetry, computer vision, 3D modeling, point cloud, multi-view, system calibration

#### ABSTRACT:

Image-based 3D models generation typically involves three stages, namely: 2D image acquisition, data processing, and 3D surface generation and editing. The availability of different easy-to-use and low-cost image acquisition solutions, combined with open-source or commercial processing tools, has democratized the 3D reconstruction and digital twin generation. But high geometric and texture fidelity on small- to medium-scale objects as well as integrated commercial system for mass 3D digitization are not available. The paper presents our effort to build such a system, i.e. a market-ready multi-camera solution and a customized reconstruction process for mass 3D digitization of small to medium objects. The system is realized as a joint work between industrial and academic partners, in order to employ the latest technologies for the needs of the market. The proposed versatile image acquisition and processing system pushes to the limits the 3D digitization pipeline combining a rigid capturing system with photogrammetric reconstruction methods.

### 1. INTRODUCTION

Nowadays image- or range-based 3D reconstruction methods are receiving a lot of attention due to the availability of fast, easy-to-use and often low-cost hardware and software solutions. 3D has proven to be a promising approach to enable precise inspection, documentation, valorization, monitoring, communication, interaction and experience. 3D digital models, often called digital twins, are increasingly used in various fields and applications, such as e-commerce, website content, heritage restoration, industrial inspection and monitoring, digital archives and cataloguing, etc.

Digital twins provide an interactive browsing experience to users that can inspect digital items by zooming-in from any viewpoint. Many applications have strong demand on the fidelity of the reconstructed geometry and texture, especially if the 3D model has to be relighted when placed into virtual showrooms.

Even though 3D technologies and processing tools for small to medium objects have been democratized (Santos et al., 2017) and approaches for semantic enrichment (Grilli and Remondino, 2019; Pierdicca et al., 2020) and access to 3D models are starting to be used, few approaches enable mass digitization of a large variety of objects, from the heritage to the industry sectors.

Controlled acquisition setups are common in the 3D field for the digitization of small to medium scale objects. These setups can include only one camera or combine cameras, structured light and laser scanners. Objects can be placed on a moving turning table and recorded from fixed sensor positions in a programmed pose sequence (Santos et al., 2014; Gattet et al., 2015; Hosseininaveh et al., 2015; Menna et al., 2017). Open-source (Stathopoulou et al., 2019) or commercial 3D reconstruction tools can then be utilized to process the collected data and derive 3D surface models.

Although solutions based on motorized turntable are low-cost and easy to implement, they often do not meet industrial requirements and constraints. The main drawback of such systems is their limited speed of digitisation given the time needed to set up the system and allow the turntable to virtually move the sensor around the object according to an acquisition schema that places the object at the centre of a sphere. In these cases, the typical acquisition protocol requires a skilled operator to set up the sensor on a stand (i.e. a tripod) and adjust its position and orientation relatively to the object according to sensor's optical characteristics (working distance, depth of field, field of view, resolution) in order to meet the project tolerances. To accomplish a complete acquisition, the operator needs to move the sensor at different heights or tilt the object relatively to the sensor. Once the set-up of the system has been carried out, the time required to turn the table to acquire all the necessary data can be as high as several minutes for a single object. Moreover, the process is significantly influenced by the operator skills due to the required human intervention and interaction. For these reasons such solutions are not well suited for systematic digitization projects in industry where productivity and consistency in the produced results is of high importance. On the other hand, multi-camera setups, integrated with motorized turn tables and linear stages, despite intrinsic instrumentation costs, have proved great productivity (Santos et al., 2014) with a number of advantages such as that they can be accurately calibrated once in a while and allow fast data acquisitions. There are also many opportunities connected to multi-camera systems for high-quality 3D digitization of objects, in particular for the benefits provided in the post-processing and editing stages with less efforts by professionals required to match market needs.

---

\* Corresponding author

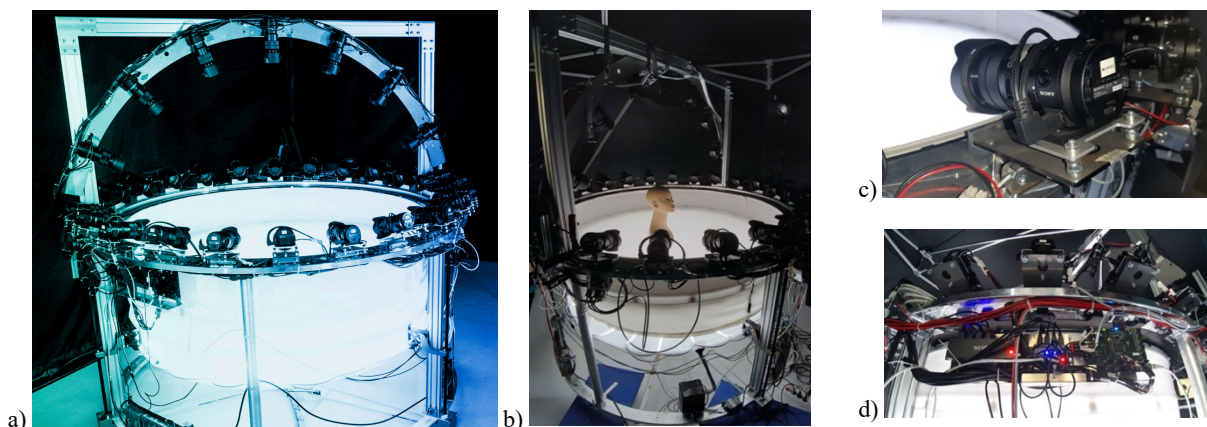


Figure 1: The DI-One multi-camera system with 31 PCB synchronized cameras, lights and a moving basement (a, b). One of the high-end professional cameras (c) and the electronics used to control groups of cameras (d).

### 1.1 Aim of the paper

In this paper we present our holistic approach for the realization of a market-ready multi-view camera system for mass 3D digitization of small to medium objects. The image acquisition and processing systems boosts the 3D digitization pipeline by jointly optimized photogrammetric reconstruction methods and a rigid capturing design and control. Compared to other solutions from the literature, the presented one (Figure 1) can deal with small up to medium objects (max size approximately  $500 \times 500 \times 500 \text{ mm}^3$ ) thanks to an adaptable FOV, it has 31 synchronized high-end professional cameras tied on a rigid structure and an adjustable moving basement.

The rest of the paper is organized as follows: Section 2 describes the realized hardware system while Section 3 presents its geometric and radiometric calibration. Section 4 presents some experiments followed by conclusions in Section 5.

## 2. MULTI-CAMERA CAPTURING SYSTEM

The Durst Imaging Product System (DI-One - Figure 1) consists of 24+7 synchronized high-end professional cameras located on a spherical ring and on a hemispherical truss (half dome) to capture lateral and upper parts of objects. A directly and indirectly lighted semi-transparent and adjustable basement allows to place objects at different heights and inclination while the cameras can capture instantly 360 degrees images for different application fields and purposes. The system comes with a proprietary PCB designed to allow the system to trigger the cameras precisely via hardware with different trigger modes. Cameras can be triggered (i) all at the same time (“freeze / standard mode”) or (ii) very accurately with a configured delay between each camera trigger (“sequential mode”) or (iii) combining the two methods (“mixed mode”). The PCB design is modular. Besides, the camera trigger the PCB also controls additional equipment of the camera system like LED / light control or trigger of an attached flash. All acquired images (a “revolution”) are transferred and written to disk each one in less than a second. A proprietary driver to control and trigger the cameras from a Linux system was developed, with a front-end application (GUI) that gives the user full control of the capturing system, i.e. change the settings of the cameras, set triggering modes, revolution history lookup, session controls, focus and zoom control, auto-focus and more.

The multi-view camera system has also a revolution playback system, i.e. a set of hardware and software components that listen for new files in a shared folder where newly created acquisitions

are stored and plays them back as either a grid of images or a rapid slideshow, mimicking the playback of the 24-image revolution as a video.

## 3. SYSTEM CALIBRATION

The system calibration includes two procedures: (i) a geometric calibration (Section 3.1) to retrieve the camera parameters and (ii) a radiometric calibration (Section 3.2) to determine the correct colours in the images and improve the visual appearance of the generated 3D models.

### 3.1 Geometric calibration

Photogrammetric self-calibration of a single or a multi-view camera system, based on ray intersection of multiple (target) points on a testfield or on a moved reference bar, was introduced by Brown (1971), Fraser (1997), Maas (1998), Gruen and Beyer (2001), Remondino and Fraser (2006).

The geometric calibration of the 31 cameras is performed using an ad-hoc testfield composed of scale bars and circular coded targets. The testfield (Figure 2) is placed in the centre of the basement in order to be visible by all cameras by rotating and tilting it in various positions. The testfield is modular and can be adjusted in size based on the necessary measurement volume. Targets of different diameters are used according to ground sample distance (GSD) and needed spatial resolution on the assets.

The calibration procedure is done in four steps:

- 1) a set of about 30 synchronized shots for all 31 cameras is acquired by tilting and rotating the testfield in the FOV of the cameras;
- 2) image orientation and bundle adjustment with self-calibration for each camera, using the 3D coordinates of the coded targets as soft constraints (coordinates are pre-measured with an estimated accuracy of better than  $15 \mu\text{m}$ );
- 3) refinement of the exterior orientation parameters through a simultaneous bundle adjustment of all the acquired images (about 900) but keeping interior parameters fixed;
- 4) average relative orientation of all the cameras with respect to a master camera.

The calibration is repeated at different zoom levels, which can then be controlled via software. Each calibration is stored in the system’s memory and utilized when needed (Fraser and Al-Ajlouni, 2006). The system accuracy and calibration verification over time is carried out using 3D length measurements of invar scale bars (Figure 2-bottom). The measurements obtained with

the DI-One system are checked against reference calibrated values (VDI/VDE, 2002).

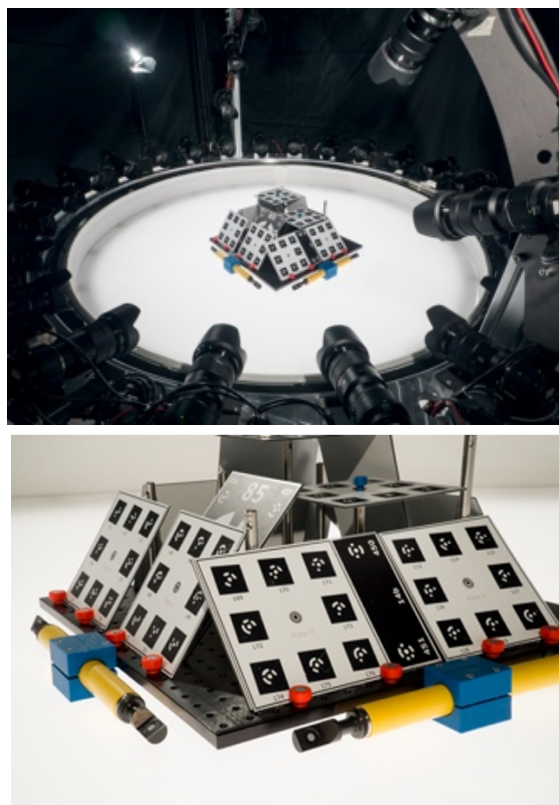


Figure 2: The used testfields for geometric calibration of the camera system (interiors and poses).

Figure 3 shows the complete image network used in the bundle adjustment (ca 900 images) together with the 3D coordinates of the testfield.

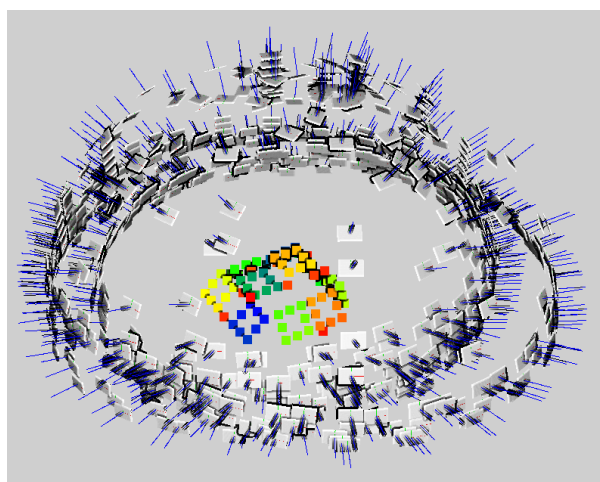


Figure 3: example of a full camera network used for the geometric calibration.

### 3.2 Radiometric calibration

The DI-One system delivers multi-camera shots in RAW image format which need to be radiometrically calibrated and post-processed to meet the colour fidelity requirements of professional applications. Assuming that the object surfaces to be acquired are

Lambertian, the acquisition process can be expressed through the equation (Gaiani et al., 2017):

$$f_c(x) = m(x) \int_{\omega} I(\lambda) \rho_c(\lambda) S(x, \lambda) d\lambda \quad \text{Eq. (1)}$$

where:

$x$  = spatial coordinate

$\lambda$  = wavelength

$\omega$  = camera's spectral domain (visible)

$c$  = colour index (R, G, B)

$f_c(x)$  = raw value of the image at position  $x$ , filter colour  $c$

$m(x)$  = Lambertian shading

$I(\lambda)$  = radiant intensity of the light source

$\rho_c(\lambda)$  = camera sensitivity function for colour  $c$

$S(x, \lambda)$  = spectral reflectance of the surface.

Equation 1 shows the dependence of a camera's pixel response on its sensitivity function  $\rho_c(\lambda)$  which generally varies from camera to camera (Pagnutti et al., 2017).

$f_c(x)$  is the raw value at position  $x$  of an image acquired with a commercial camera, and in a commercial imaging pipeline it undergoes multiple steps of post-processing, such as Bayer interpolation, noise subtraction, and white balancing.

If a faithful chromatic reproduction of an object is required, these values, which represent the camera's own RGB values, must be linked to a device-independent colour space, such as CIEXYZ or CIELAB. This process is referred to as the radiometric camera's calibration (Westland et al., 2012) and it can be performed in various ways.

The most widely used characterization approaches are described in the international standard ISO 17321-1 (ISO17321): one is based on the measurement of the cameras' spectral sensitivities, while the other is target-based.

We performed the latter procedure for each camera in the multi-camera system. Since the illumination is controlled, the colour calibration was performed for a D50 standard illuminant.

We printed an IT8.7/4 colour chart with 1617 patches on a ProofMaster Matt 140g paper (Fig. 4). Although such a target has been designed for the colour characterisation of scanners, monitors and other output devices, it has been previously exploited to perform multi-camera calibration (Troester et al, 2018).



Figure 4: Example of captured IT8.7/4 colour chart

We then measured the CIELAB coordinates of the target with a Barbieri Electronic LFPqb spectro-photometer (Barbieri, 2020), and we used them as references.

The colour chart was taped on a flat white matte holder and was acquired with each camera of the system at a position roughly normal to its optical axis. A spotlight with 5000 K colour



temperature was placed at about 45 degrees with respect to the camera axis. The other acquisition parameters correspond to those previously discussed.

On each image we applied the Python's module rawpy (<https://pypi.org/project/rawpy>) for an AHD demosaicing and for the extraction of the raw RGB values of each colour patch. It is then possible to link these values to the previously measured CIELAB values in order to obtain a proper colour profile, which can then be used to correct the texture file of a 3D model.

In particular, we used a multi-dimensional look-up table defined on a  $33 \times 33 \times 33$  grid in the raw RGB space of each camera (Balasubramanian and Klassen, 2003), implemented using the Python's scikit-learn tool (<https://scikit-learn.org/>).

An example of the radiometric processing is shown in Fig. 5. The improvement in the appearance of the scene is evident. If necessary, further corrections could be performed, such as bad pixel removal, bias and dark frame subtraction, flat fielding, etc.

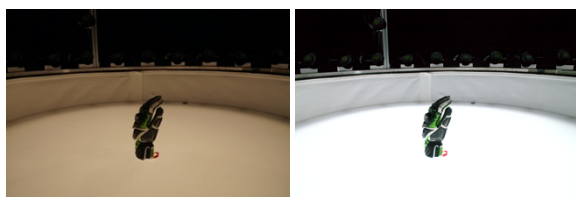


Figure 5: Example of the radiometric calibration result (right) on a raw image (left) of a glove.

#### 4. 3D RECONSTRUCTION PIPELINE

The DI-One system (<https://www.covisionlab.com/media-lab>) is coupled with an automated photogrammetric pipeline which allows to generate textured 3D models of various objects (Figure 6). Using the camera calibration parameters, the images are remapped as if they were all obtained through the same pinhole camera, i.e. without geometric distortions, with a unique principal distance (focal length) and the principal point at the center of the image format. This procedure simplifies the 3D reconstruction and modelling pipeline and allows the use of the DI-One images also with other research and commercial based software, which might not be able to deal with distortion parameters or use a different distortion formulation.

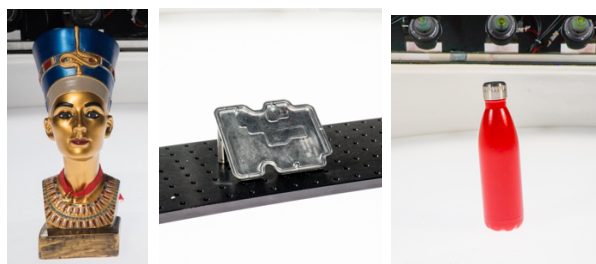


Figure 6: Examples of some objects digitized with the realized DI-ONE multi-camera system.

For objects characterised by cooperative texture and simple geometry, a single DI-One synchronized shot (31 images) may be sufficient for reconstructing their visible parts. In these cases, the exterior orientation parameters (camera poses) obtained from system calibration (Section 3.1) are directly used. Figure 7 shows the coordinate system and the camera network as defined from the calibration stage. These exterior orientation parameters are directly provided to the photogrammetric pipeline for the 3D reconstruction procedure, thus providing also the proper scaling.

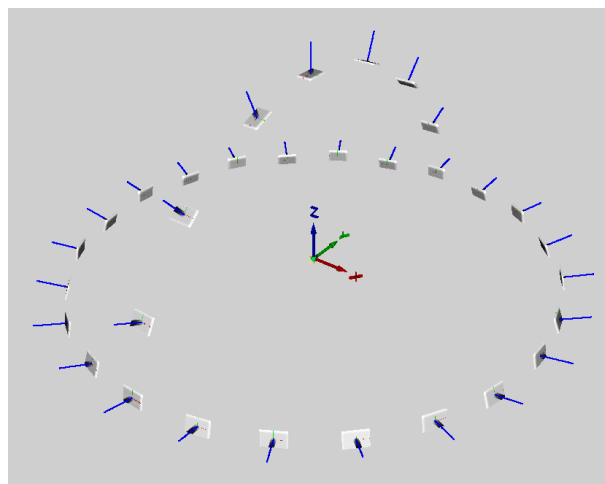


Figure 7: DI-One camera network and coordinate system.

The typical pipeline starts with a background masking, performed automatically through image subtraction procedures (Figure 8) and then feature points are extracted and matched within the unmasked areas of the images. These tie points are then triangulated using the exterior orientation parameters and the individual interior orientation parameters derived from the calibration procedure. Dense image matching procedures are then applied to generate depth maps and a dense cloud. Finally, a mesh is triangulated over the point cloud and textured.



Figure 8: Example of image masking obtained through background subtraction.

Figure 9 shows an example of a 3D reconstruction of small vase ( $170 \times 130 \times 180 \text{ mm}^3$ ) obtained with 31 synchronized images acquired under the camera network shown in Figure 7.



Figure 9: One (out of 31) synchronized image of a small vase obtained with the DI-One system (left) and the reconstructed 3D mesh (right).

For more complex objects (such as the glove in Figure 12), depending on their shape, self-occlusions and texture characteristics (shiny, texture less, etc), more shots may be required after tilting or rotating the object within the field of view of the system. Different camera networks can thus be obtained in order to achieve a complete 3D reconstruction. Figures 10 and 11 show two examples of camera networks obtained with a total of 4 DI-One shots by rotating the object respectively around the Z



axis by the angles  $\{-48.75, 48.75, 97.5\}$  [deg] (Figure 10) or by tilting the object around the Y axis of the angles  $\{-22.5, 22.5, 45\}$  [deg] (Figure 11).

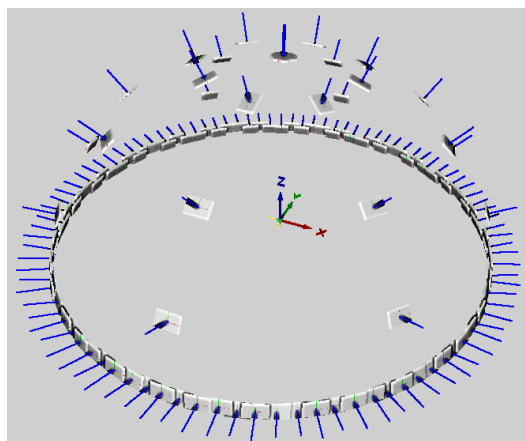


Figure 10: A typical camera network obtained rotating the object around the Z axis in order to have a better coverage.

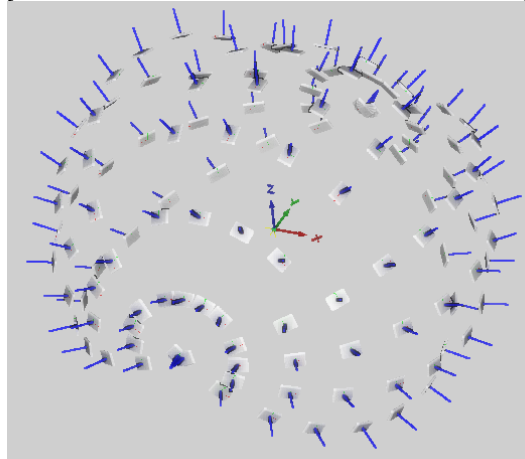


Figure 11: Examples of camera network obtained using multiple shots after tilting the object around the Y axes.

Figure 12 shows two rendered views of the 3D texturized model of a Durst Gil camera (production 1938), reconstructed from two shots to capture the entire object texture and geometry.



Figure 12: Views of a 3D model of the 1938 Durst Gil camera reconstructed from two shots taken at different object poses.

Figure 13 shows the 3D reconstruction of a skiing glove pictured with three synchronized shots (93 images) and textured with a radiometrically corrected texture.



Figure 13: One (out of 93) image of a glove obtained by rotating the system basement around the Z axis (left) and the reconstructed 3D mesh through photogrammetric process (right).

## 5. CONCLUSIONS

The paper presented the realization of the Covision Media Lab (<https://www.covisionlab.com/media-lab>) 3D digitization process named DI-One multi-camera system. DI-One is a market-ready image-based solution composed of 31 synchronized high-end professional cameras for mass 3D digitization of small to medium objects. The system and processing methodology were conceived as a joint work between industrial and academic partners, they are quite versatile and push to the limits the 3D digitization by combining a rigid capturing system with photogrammetric reconstruction methods.

## ACKNOWLEDGEMENTS

The authors would like to highlight the contribution of Durst Phototechnik AG in this effort to manufacture the Product System DI-One.

## REFERENCES

- Balasubramanian, R., Klassen, R. V., 2003. Efficient color transformation implementation. *Digital Color Imaging Handbook*, CRC Press: Boca Raton, FL, USA.
- Barbieri, 2000. <https://www.barbierielectronic.com/en/digital-imaging/1-0.html> (last access: May, 2020).
- Brown, D., 1971. Close-Range Camera Calibration. *Photogrammetric Engineering*, Vol. 37, No. 8, pp. 855-866.
- Fraser, C.S. and Al-Ajlouni, S., 2006. Zoom-dependent camera calibration in digital close-range photogrammetry. *Photogrammetric Engineering & Remote Sensing*, 72(9), pp.1017-1026.
- Fraser, C.S., 1997. Digital camera self-calibration. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 52, pp. 149-159.
- Gaiani, M., Apollonio, F. I., Ballabeni, A., Remondino, F., 2017. Securing color fidelity in 3D architectural heritage scenarios. *Sensors*, 17(11), 2437.
- Gattet, E., Devogelaere, J., Raffin, R., Bergerot, L., Daniel, M., Jockey, Ph., De Luca, L., 2015. A versatile and low-cost 3D acquisition and processing pipeline for collecting mass of

archaeological findings on the field. *ISPRS Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. XL-5/W4.

Grilli, E., Remondino, F., 2019. Classification of 3D Digital Heritage. *Remote Sensing*, Vol. 11(7), 847.

Gruen, A. and Beyer, H.A., 2001. System calibration through self-calibration. In 'Calibration and Orientation of Cameras in Computer Vision' Gruen and Huang (Eds.), Springer Series in Information Sciences 34, pp. 163-194.

Hosseiniaveh, A., Yazdan, R., Karami, A., Moradi, M. and Ghorbani, F., 2015. A low-cost and portable system for 3D reconstruction of texture-less objects. *ISPRS Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. 40(1/W5), p.327.

ISO 17321-1:2012, 2012. Graphic technology and photography—Colour characterisation of digital still cameras (DSCs) - Part 1: Stimuli, metrology and test procedures.

Menna, F., Nocerino, E., Morabito, D., Farella, E. M., Perini, M., Remondino, 2017. An open source low-cost automatic system for image-based 3D digitization. *ISPRS Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. XLII-2/W8, pp. 155-162.

Pagnutti, M. A., Ryan, R. E., Cazenavette, G. J., Gold, M. J., Harlan, R., Leggett, E., Pagnutti, J. F., 2017. Laying the foundation to use Raspberry Pi 3 V2 camera module imagery for scientific and engineering purposes. *Journal of Electronic Imaging*, Vol. 26(1), 013014.

Pierdicca, R., Paolanti, M., Matrone, F., Martini, M., Morbidoni, C., Malinverni, E.S., Frontoni, E., Lingua, A.M., 2000. Point Cloud semantic segmentation using a deep learning framework for Cultural Heritage. *Remote Sensing*, 12, 1005.

Remondino, F., Fraser, C., 2006. Digital camera calibration methods: considerations and comparisons. *ISPRS Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. 36(5), pp. 266-272.

Santos, P., Ritz, M., Tausch, R., Schmedt, H., Monroy, R., De Stefano, A., Posniak, O., Fuhrmann, C., Fellner, D.W., 2014. CultLab3D - On the verge of 3D mass digitization. *Proc. Eurographics Workshop on Graphics and Cultural Heritage*.

Santos, P., Ritz, M., Fuhrmann, C. and Fellner, D., 2017. 3D mass digitization: a milestone for archaeological documentation. *Virtual Archaeology Review*, Vol. 8(1).

Stathopoulou, E.-K., Welpner, M., and Remondino, F., 2019. Open-source image-based 3D reconstruction pipelines: review, comparison and evaluation. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W17, 331-338.

Troester, P., Nestler, R., Beger, M., Urban, P., 2018. Fogra-Forschungsbericht Nr. 11.002 Prozessstandardisierung im grafischen 3D-Druck fuer 3DP- und FDM-Verfahren, Munich, Germany. Available at: <https://fogra.org/en/fogra-research/prepress/research-projects-prepress/graphical-3d-printing/publications-3d.html> (last access: May, 2020).

VDI/VDE, 2002. Guideline 2634 part 1, Optical 3D measuring systems - Imaging systems with point-by-point probing. VDI/VDE-Gesellschaft Mess-und Automatisierungstechnik.

Westland, S., Ripamonti, C., Cheung, V., 2012. *Computational colour science using MATLAB*. John Wiley & Sons.