Journal of Archaeological Science 66 (2016) 137–145

Contents lists available at ScienceDirect

Journal of Archaeological Science

journal homepage: http://www.elsevier.com/locate/jas

Accurate measurement with photogrammetry at large sites

P. Sapirstein

Department of Art & Art History, Center for Digital Research in the Humanities, University of Nebraska–Lincoln, 120 Richards Hall, P.O. Box 880114, Lincoln, NE 68588-0114, USA

ARTICLE INFO

Article history:
Received 21 May 2015
Received in revised form
16 November 2015
Accepted 3 January 2016
Available online 21 January 2016

Keywords:
3D recording
Photogrammetry
Accuracy
Best practice
Historical architecture
Digital heritage
Olympia

ABSTRACT

Photogrammetry has become increasingly popular as a low-cost method for documenting cultural heritage and archaeological excavations. However, we have yet to establish best practices for its implementation at the site, or methods for assessing the accuracy of the resulting 3D measurements. This article presents a recent study of the Temple of Hera at Olympia, where a 25 × 55 m area was recorded at 1 mm resolution using photogrammetry both for survey and 3D reconstruction. Coded targets were set up throughout the site, which was then photographed in two phases. First, a site-wide survey established the locations of the network of targets. Second, sets of close-up photographs for detailed 3D reconstruction of the site were registered to the global survey via the targets. This technique developed at Olympia improves measurement accuracy by an order of magnitude compared to previous implementations, with a precision of at least 1 part in 50,000, and 95% of the surfaces located accurately within 2–3 mm.

1. Introduction

Photogrammetry is all the rage these days in archaeological fieldwork. Many projects have already implemented a complete 3D recording system based on the technology (e.g. De Reu et al. 2013; De Reu et al. 2014; Dellepiane et al. 2013; Fernández-Hernandez et al. 2014; Olson et al. 2013; Roosevelt et al. 2015; Stal et al. 2014). Modern photogrammetric systems are largely automated, using Scale Invariant Feature Transform (SIFT), Structure from Motion (SfM), and Multi-View Stereo (MVS) to restore complex 3D geometry from sets of photographs (Luhmann, 2010; Vergauwen and Van Gool, 2006). The software can rapidly create high-resolution colour 3D models of an on-going excavation or a building site at much lower cost than laser-scanning hardware. Photogrammetry now has the potential to revolutionise how archaeologists document, study, and preserve antiquity.

If we are going to adopt this technology for digital recording, we should also discuss how it is implemented at the site. How should the job be set up and processed within the software? How accurate are the estimated camera positions, points, and surfaces? Massive failures are possible when the photographs do not overlap sufficiently, and movement within the scene during photography has unpredictable consequences. Yet with the right photographs, the software is able to create detailed 3D models that look convincingly lifelike. The beguiling realism makes it all the more critical to examine the accuracy of the results. If we are to determine best practices for photogrammetry in archaeology, assessing accuracy is essential for comparison of different implementations. As one paper recently published in this journal concluded, “until structure from motion can demonstrate reliable accuracy, and this can be calculated on a case by case basis, it is unlikely to be taken seriously as a measurement tool.” (Green et al. 2014, p. 181).

2. Previous research

The question of accuracy is difficult to address directly, because the extensive automation of SfM/MVS software makes its operation essentially a “black box”. One approach has been to test error of individual measurements produced by SfM, typically by comparison to a set of reference points measured with a Total Station. Examples are compiled in Table 1a.

For each project, an estimate of the *precision* has been expressed as a proportion 1:*k*, where *k* is the size of the scene divided by the reported standard error (Fraser and Brown, 1986). This metric has no inherent scale. A hypothetical camera system with a 1:5000 precision could distinguish measurements down to 1 mm across a 5-m-long vehicle, but only to 1 m when used to measure aerial

E-mail address: orientalizing@gmail.com.

http://dx.doi.org/10.1016/j.jas.2016.01.002
0305-4403/

Table 1
Photogrammetric errors reported in recent studies.

Study	Subject	Scale (m)	Error (mm)	Precision (1-σ)
a: Ground Control Point measurement errors reported in SfM-based surveys				
Reinoso et al. 2014	Buildings	10–20	10–26	<1:1000
Olson et al. 2013	Trenches	5–35	~20–40	1:500
Remondino et al. 2012	Buildings	~15	35	1:400
Green et al. 2014	Buildings	5–15	19–39	<1:400
De Reu et al. 2014	Trenches	~10	8–15	1:1000
Koutsoudis et al. 2014	Building	10	14	1:700
Riveiro et al. 2011	Building	<10	12	1:800
De Reu et al. 2013	Trenches	2.5–6	9–25	1:500
Dai and Lu 2010	Object	<2.5	2	1:1000*
Dellepiane et al. 2013	Trenches	~2.0	2	1:1000*
Martínez et al. 2015	Pavement	~0.6	2	1:300*
b: Errors in vertex positions of meshes created by MVS				
Doneus et al. 2011	Trenches	~10	18	<1:600
Koutsoudis et al. 2014	Building	~10	14	1:700
Dellepiane et al. 2013	Trenches	~2.0	6	>1:300*
Remondino et al. 2008	Building	~1-2	1.4	<1:1500
Remondino et al. 2009	Building	<1.2	<0.4	1:3000
Lerma and Muir 2014	Object	1.0	0.2	1:5000
Jennings and Black 2012	Objects	0.5	0.2	1:2500*
Kersten and Lindstaedt 2012	Objects	0.5	0.3	>1:1500
Koutsoudis et al. 2013	Object	0.2	0.07	>1:2500

a: All studies in the first group used a Total Station to measure control points to establish SfM errors, except those starred (*) used a tape measure or callipers to check distances between two points measured by SfM. Normally error is reported as the standard deviation (1-σ) or RMS.
b: All studies in the second group used laser hardware as the reference for quantifying error in the MVS-generated vertices, except (*) Jennings and Black 2012 used 3D-printed models of known geometry, and Dellepiane et al. 2013 used repeatability tests—comparing models of the same area from separate batches of photographs.

photographs of a 5 km region. The 1:*k* proportion is only a rough estimate of the true precision, which will vary within every image (Barazzetti et al. 2011; Fraser et al. 2005).

While the absolute errors in Table 1a vary considerably (2–40 mm), the precisions are fairly consistent, despite very different subjects, conditions, and equipment. Other studies have instead examined the positional accuracy of the vertices in the 3D mesh derived by MVS. Usually the photogrammetric models have been compared to a laser scan of the same subject (Table 1b). The estimated precision of the mesh vertices is higher, though all of the cases with precision greater than 1:1000 involve small objects or segments of walls, and were performed under more carefully controlled conditions than the field surveys in Table 1a. Nonetheless, we should have expected the automatically generated vertices of the full 3D meshes to contain *higher* errors than the control points measured with SfM, because in MVS detailed surfaces are reconstructed from a scene structure determined by SfM.

Regardless, these results are more than an order of magnitude below the accuracy achieved with close-range photogrammetry since the 1970s. In close-range applications, the points are measured from machine-readable coded targets in the scene. A large-format "metric" camera can be calibrated to measure targets at precisions up to 1:300,000–1,000,000 (Fraser, 1992; Luhmann, 2010). Modern digital cameras lack the internal stability of metric cameras, but an SLR can still be calibrated for precisions as high as 1:50,000–100,000 (Fraser and Al-Ajlouni, 2006; Galantucci et al. 2014; Luhmann, 2010; Rieke-Zapp et al. 2009; Stamatopoulos and Fraser, 2011; Zhenzhong et al. 2010). The high performance was established using coded targets. While it is possible to use the large numbers of feature points automatically detected through SIFT for calibration instead of targets, measurement is somewhat less precise, in the range of 1:5000–20,000 (Barazzetti et al. 2011; Stamatopoulos and Fraser, 2014).

The tables above suggest that recent applications of SfM-based photogrammetry have discarded a key strength of close-range photogrammetry: its potential for extraordinarily high precision. Can we retain this high precision while using SIFT/SfM/MVS?

This question was addressed during recent photogrammetric recording at the site of Olympia, part of the Digital Architecture Project. The Hera temple is one of the best-preserved early Doric temples surviving from the Greek world (Dörpfeld and Schleif, 1935). Its walls are preserved to a metre above the ground, and most of its 40 peristyle columns are partially intact (Fig. 1). The 55 × 25 m building site includes several columns standing almost 7 m above the foundations. The study of the extensive and spatially complex remains of the Heraion requires the sort of detailed recording for which photogrammetry is ideal (Sapirstein, 2015). The photogrammetric survey at Olympia was also designed to achieve the highest accuracy possible, within the temporal and budgetary constraints typical of archaeological fieldwork. The system is a step toward best practices for outdoor, large-scale applications of SfM/MVS photogrammetry.

3. Methods for quantifying error

Error is presented here as both *precision* and *accuracy*. Precision indicates the finest measurement possible and is represented by the RMS of the discrepancies of 3D points from a set of reference measurements. Accuracy indicates the expected error and is estimated at the 2-σ confidence level (CL). The RMS and the 95.5% CL are relatively simple to calculate and are popular in the literature (e.g., Dai and Lu, 2010; Höhle and Höhle, 2009; Luhmann, 2010).

The challenge is to establish the set of reference coordinates from which to quantify these values. Typically a Total Station is used for point measurements at buildings or trenches, but the error of this hardware will usually exceed that of photogrammetry (Toschi et al. 2014). In field conditions, it is difficult to reach a 2-σ CL below ±5–10 mm, due to the ±2 mm error in distance measurements, error in the set up of the station, the position of the reflector, and small movements of the station during survey (Sapirstein, 2015). Assuming a relatively low photogrammetric precision of 1:10,000, it should be possible to measure features down to 1 mm within a 10 m scene, an order of magnitude below the error of the Total Station. In fact, the low precisions of 1:1000 or less reported in Table 1a are probably due to error in the Total Station survey rather than the photogrammetry.

Fig. 1. View of the Hera temple from the east. Credit: https://www.flickr.com/photos/letdown102/1535276795/ (Creative Commons license 2.0, attribution, noncommercial, noderivatives).

However, equipment capable of high-precision measurements of 3D points like a CMM is ill-suited for the conditions of an archaeological site, leaving few options for determining the reference points. The first approach adopted here was testing the equipment used at Olympia against a metric camera with demonstrably higher precision. The second has been *repeatability* tests, or comparisons of different batches of photographs to a reference set of coordinates established from all the photographs combined within a single calibration. The latter approach lacks reference points established by higher-precision machinery, but it has the advantage that it can be implemented at any site. Repeatability tests have been standard in close-range photogrammetry testing when no more precise measuring technique is possible (e.g. Fraser, 1992; Stamatopoulos and Fraser, 2014). Repeatability tests are also used to assess the quality of the 3D meshes generated of the Hera temple (as in Dellepiane et al. 2013; Galantucci et al. 2014). Otherwise, archaeologists would have to bring expensive laser-scanning hardware to the site every time they tested the accuracy of their data, which would mitigate the reason for using photogrammetry in the first place: its low cost and efficiency.

The following discussion is organised by the stages of photogrammetric processing, each of which contributes error:

(1) Camera: with perfectly exposed and focused images, the stability of the lens projection determines the theoretical maximum precision and accuracy.
(2) Orientation: the parameters for each photograph describing the position and lens calibration. The relative orientations of every camera are estimated through bundle adjustment, which may contribute additional error above the theoretical precision attainable in a single image.
(3) Dense point extraction: the software reconstructs depth throughout the imagery (MVS), which is prone to additional error.
(4) Meshing/texturing: 3D meshes approximating the surfaces are interpolated from the dense point cloud, and textures are applied by re-projecting the photographs onto the resulting surfaces.

3.1. Camera precision

The camera used at Olympia was a Nikon D800e, with 7360 × 4912 pixels over a 35.9 × 24 mm sensor, paired with a manual focus 28-mm lens (Nikkor 28 mm f/2.8 AI-s). A direct assessment of precision below the pixel resolution of the sensor (ca. 1:7500) would be difficult from points marked manually or generated automatically by SIFT, which detects features at the level of a pixel (Lowe, 2004). Precision is better tested with coded targets, whose centres can be estimated precisely well below 1/10th of a pixel (Fraser et al. 2005; Luhmann, 2010).

First, the accuracy of the D800e was tested by comparison to an Alpa 12 metric camera with a fixed-focus Schneider 35 mm f/5.6 Apo-Digitar XL lens, and an Aptus-II/6 digital back (44 × 33 mm sensor). The systems were tested outdoors with 50 targets set in a 4 × 3 × 2 m area. Testing was with Agisoft PhotoScan Pro, a commercial SfM/MVS package supporting coded targets (Agisoft PhotoScan Pro manual). Both cameras were calibrated from 60 images each, and the reference coordinates determined for 38 targets detected by the software and a separate scale bar. Five of the 38 were used as ground control points to set a common coordinate system, and the others used for testing. For both cameras, 4 batches of 10 photographs each were separated for internal repeatability tests.

The metric camera significantly outperformed the D800e used at Olympia. Its internal repeatability had an RMS of 0.025 and a 2-σ CL of 0.047 mm, suggesting a precision of about 1:160,000 (calculated by dividing the 4000-mm width of the scene by the RMS). The values for the D800e were 0.070 and 0.116 mm, respectively, when compared to the reference points of the metric camera, suggesting a minimum precision of ca. 1:50,000.

3.1.1. Orientation

A similar outdoor scene in Greece was used to test the effect of different program settings on the accuracy of the orientation (Fig. 2). The tests examined the inclusion of coded targets and different approaches to calibration. First, the reference coordinates were determined from 275 photographs to optimize the lens calibration, taken in 12 sets of 20–25 photographs. Several targets had to be discarded after movement during photography, leaving 40 for analysis, of which five were reserved as ground control points. Individual sets were then tested against the 35 remaining reference points.

First, in the absence of a more precise system to establish reference coordinates, it is encouraging that the internal repeatability tests shown in the first column of the table (Table 2.1) confirmed the 1:50,000 precision already determined for the D800e against the metric camera. Error was only slightly higher when calibrating each set of photographs separately (Table 2.2). These tests also demonstrate a superior accuracy by including targets. Target-based calibration has recently been questioned by Stamatopoulos and Fraser (2014), who found in a similar testing scenario their calibration from SIFT-generated tiepoints was slightly more precise than calibration based exclusively on coded targets. The difference in the testing at Olympia presented here is that *both* the many automatically detected tiepoints and the precisely located targets were used for calibration, raising accuracy by at least 70% compared to excluding the targets (Table 2.2 vs. 2.3).

Though a fixed lens is advisable for photogrammetry, many archaeologists have used zoom lenses instead (e.g. De Reu et al. 2013; De Reu et al. 2014; Doneus et al. 2011; Kersten and Lindstaedt, 2012; Koutsoudis et al. 2013, 2014; Lerma and Muir, 2014; Olson et al. 2013). Zoom lenses usually include features like optical stabilisation that actually destabilise the lens projection, potentially reducing the measurement accuracy (Fraser, 2012; Toschi et al. 2014).

We can simulate the effect of a zoom lens by setting the software to vary the calibration for every image, as if they had been taken with a variable lens (Table 2.4, 2.5). The accuracy is degraded significantly, though the fourth target-based scenario is significantly better than the target-less one. This was also tested directly by photographing the same scene with a zoom lens (Nikkor 18–300 mm VR) set at 28 mm with the autofocus and image-stabilisation systems allowed to alter the projection in each photograph. Because this lens is incompatible with the D800e, the tests instead use a Nikon D7100 with a 23.5 × 15.6 mm (6000 × 4000 pixel) sensor (Table 3).

As before, the coded targets are valuable. Leaving them out of the per-image calibration increases error by about 45% (Table 3.1 vs. 3.2). Including targets allows a per-image calibration that corrects

Table 2
Observed error (mm) for the D800e with a fixed 28-mm lens.

Calibration	1. Precal.	2. Single	3. Single	4. Vary	5. Vary
Targets	Yes	Yes	No	Yes	No
n	412	412	140	140	140
RMS	0.09	0.11	0.19	0.27	0.50
95.5% <	0.15	0.19	0.31	0.45	0.92
Precision	1:48,000	1:36,000	1:21,000	1:14,000	1:8000

Calibration: precal. = a single calibration used amongst all batches of photographs; single = a new calibration for each batch of photos; *vary* = different calibrations for each photograph.
Targets: yes = calibration including the coded targets; no = calibration using only the SIFT-generated feature points.

Table 3
Observed error (mm) for a D7100 with zoom at 28 mm, image stabilisation, and autofocus (n = 96). Conventions as in Table 2.

Approach	1. Vary	2. Vary	3. Single	4. Single
Targets	Yes	No	Yes	No
RMS	0.21	0.31	0.54	0.63
95.5% <	0.37	0.53	1.01	1.23
Precision	1:19,000	1:13,000	1:7000	1:6000

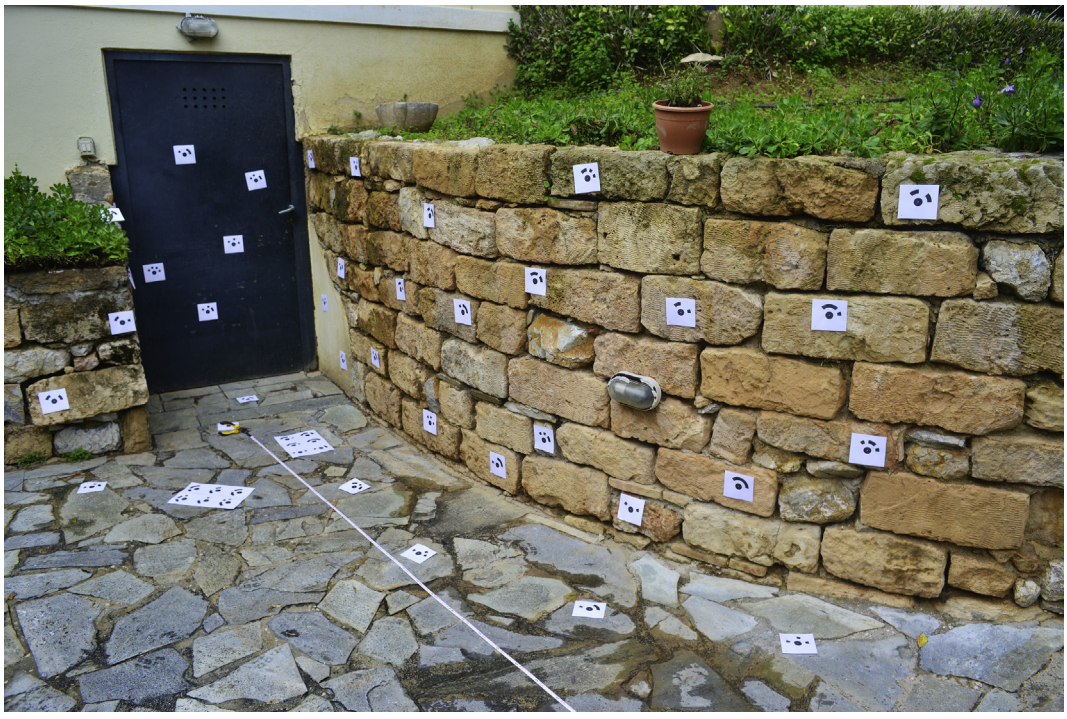

Fig. 2. Outdoor test scene in Greece. Credit: author.

most variations in focus and centring. In fact, the fixed lens outperforms the variable zoom lens by surprisingly little in the individually calibrated sets (Table 3.1 vs. 2.2). The lesser dimensions and resolution of the D7100's sensor can account for most of the reduced accuracy, and in fact the D7100 outperformed the D800e when the calibration was allowed to vary (Table 3.1 vs. 2.4). This can probably be attributed to the D7100's smaller sensor, which crops the perimeter of the image projected by the lens. The wider field captured by the larger sensor of the D800e suffers from more blurring and chromatic aberrations near its borders.

The final columns of Table 3 show a common problem built into photogrammetric software. If every photo imported is taken at the same zoom setting, the program will by default calculate a single calibration for all the images. Although the correct behaviour for a fixed lens, this would be a serious mistake with the variable zoom lens. With instable settings during photography, the calibration *must* be allowed to vary or else error increases greatly—almost tripled in this case, regardless of whether targets are included or not (Tables 3.3–3.4).

These tests make clear that machine-readable targets should be included in any photogrammetric survey whose purpose is not just to produce an aesthetically appealing model, but also an accurate one. The targets also greatly speed up processing because the user does not need to mark ground control points manually (Roosevelt et al. 2015). More importantly, it would be difficult to assess the accuracy of the photogrammetric survey without them in the scene. While target-less calibration to about 1:20,000 is possible with the equipment tested here, it would be hard to discern this without the targets, which are the only features in the scene that can be easily located at sub-pixel precision. Indeed, previous studies advocating for target-less calibration included targets to assess their target-less surveys (Barazzetti et al. 2011; Stamatopoulos and Fraser, 2014).

3.1.2. Orientation at Olympia

These tests suggest that, at Olympia, the equipment should be able to locate coded targets within 2 mm over the 55-m building site—up to a five-fold improvement over a Total Station survey (±5–10 mm). However, the site is vastly more complex than the testing scenarios. The camera would have to be elevated more than 40 m above the building to capture the whole site in one frame, but at that distance the architecture would be indistinct. A sub-millimetre resolution is needed in order to capture details such as the joint seams between adjacent blocks, so photographs must be taken at 2.0–3.5 m distance (Sapirstein, 2015).

A two-stage survey was implemented to meet these demands. First, an aerial survey established the ground control points. A network of 300 targets set throughout the site was photographed with the camera raised on a boom (Fig. 3). At a ca. 9 m elevation, the camera captures only about 10% of the site, but 500 overlapping photographs ensured full coverage at adequate detail to detect the coded targets. Second, close-up photographs were taken in batches. These allow a more detailed reconstruction of the surfaces by MVS, while registering the batches within the coordinate system determined in the aerial survey. Altogether there were 17 batches with between 50 and 500 close-up photographs for various segments of the building.

During the initial aerial survey, the target coordinates needed to be set within the Olympia site grid. A network of 85 iron stakes marked with a 1-mm dot were surveyed with a Total Station and manually marked in the aerial photographs, registering the aerial survey to the site coordinates. The survey was also constrained by several stake-to-stake distance measurements, including precise lengths taken between nearby stakes, and several longer distances checked against both a 30-m metal tape ruled at 1 cm intervals, and

Fig. 3. Camera raised on a boom. Credit: Dr. Sarah C. Murray.

an older survey of the building (Sapirstein, 2015). The quality of the aerial survey is assessed in Table 4:

First, the photogrammetry refined the input coordinates for the stakes measured by the Total Station and tapes. Including all the nails that could be seen in at least 3 aerial photographs, and removing 4 outliers that had moved during the survey, left 82 points whose RMS was almost 7 mm (Table 4.1). The photogrammetric survey was far more internally consistent. This was tested by comparing the aerial survey coordinates to those determined in the batches of close-up photographs, which were several times more precise than the aerial survey and thus a good set of reference measurements. Because many points appeared in several batches of close-up photographs, there were 765 comparisons in all. Of these, 95% were within 1.3 mm of the aerial survey (Table 4.2). Furthermore, after removing indistinct targets in the close-up sets and the stakes—which cannot be marked as accurately as the targets—the 2-σ CL dropped to just 1.0 mm (Table 4.3). This accuracy is higher than the 1.3–1.9 mm predicted from the initial tests of the equipment (above, Table 2.1). The aerial survey at Olympia may have outperformed the test scenarios because of better illumination, the greater distance from the subject (which extends the depth of field and clarity of the images), and the relatively large number of photographs in the aerial survey (see Fraser et al. 2005).

The table also reiterates the importance of the coded targets. Several close-up sets were recalibrated using only the SIFT-generated tiepoints, more than doubling error (Table 4.4 vs. 4.5). Had a variable zoom lens been used with a single calibration for all the photos (as in Table 3.3, and common in archaeological

Table 4
Discrepancies from the photogrammetric aerial survey (mm).

	1. Total station: Stakes	2. Close-up sets: all points	3. Close-up sets: just targets	4. Close-up sets: all points[a]	5. Close-up sets: all pts, *SIFT*[a]
n	82	765	507	662	592
RMS	6.8	0.73	0.55	1.46	3.16
95.5%<	11.1	1.29	0.98	1.91	5.01
Precision	–	1:68,000	1:91,000	1:34,000	1:16,000

[a] Some points not visible in the aerial survey or which shifted during the work are excluded from the previous columns, but included in these calculations.

photogrammetry), the accuracy would drop below 1:3,500, approaching a ±20 mm error unacceptable for architectural study.

4. Error in the 3D models

Next to consider is the quality of the dense point clouds and meshes created by MVS from the oriented and calibrated photographs. Because of the varying levels of detail during the survey, there are two different levels of potential error:

Global error: the reliability of measurements across the site from the initial survey.

Local error: the reliability of surfaces restored from close-up photographs.

The local error includes millions of vertices whose quality may be much lower than that of the initial tiepoints. After MVS, it is not uncommon for a few points to be reconstructed well outside the actual surfaces, which are easily identified and deleted by the operator. The local error that remains in the cleaned point cloud and meshed surfaces will be affected by:

- Poor lens calibration and initial orientations
- Blurring in the photographs
- A lack of high-contrast patterns on the surfaces, or an excess of thin, elongated objects like vegetation
- Surfaces oblique to the camera, such as crevices at the joint seams between blocks
- Movements in the scene during photography

On these counts, Olympia is relatively well suited for accurate dense surface modelling. However, since natural sunlight was the only illumination, the lighting and shadows changed throughout the day, limiting photography to brief windows of time.

4.1. Dense point extraction

Without laser-scanning hardware, local error at Olympia was estimated with repeatability tests. Models created from separate batches of photographs of the same areas of the temple were compared. Table 5 presents the 3D point discrepancies estimated using CloudCompare (CloudCompare manual). The precision, calculated by dividing the mean error into the 2.5-m width captured in a typical photograph, is only a rough estimate and included primarily for comparison to previous studies (Table 1b).

Table 5
Dense point cloud and mesh discrepancies at Olympia (mm).

Type	1. Cloud	2. Mesh	3. Mesh	4. Mesh
Area	NW	NW	Walls	Corners
(vertices) *n*	36Ma	1.3Ma	1.3Ma	600k
mean	0.31	0.49	0.85	1.25
95.5%<	0.90	1.43	3.03	5.64
(3-σ) 99.7%<	2.31	9.41	16.5	16.9
Precision	1:8000	1:5000	1:3000	1:2000

NW = north wall, comparisons of two ca. 1.5-m square areas with Full ("Ultra") quality modelling.
Walls = two 20-m areas of the cella walls, at 1/16th ("Medium") subsampling.
Corners = a ca. 2 × 2 × 1 m volume at the NE and SW corner of the peristyle, 1/16th subsampling.

First are the interpolated points from MVS. Three clouds of 15–25 million points each were generated using every pixel in the photographs ("Ultra" quality in PhotoScan) for two comparisons at the northwest area of the cella (Fig. 4a). Table 5.1 shows results are repeatable within the millimetre at the 95.5% CL. However, this method somewhat exaggerates the error compared to the calibration tests because there is no higher-quality reference data for comparison. If we assume the true surface was somewhere between the two compared sets of points, the discrepancies would be lower. The crevices in the joint faces between adjacent blocks of the masonry were impossible to capture effectively, and produced almost all the 3-σ discrepancies (the top 0.3% of the distribution) of 2–5 mm (Fig. 4). On the other hand, points in clearly visible areas should be plotted more accurately, perhaps within 0.5 mm at the 2-σ CL. With full image sampling, points can be restored accurately up to the equivalent of one pixel in the camera sensor.

The roughly 1:8000 precision at Olympia somewhat exceeds those in previous reports (Table 1b). This may be due to the precision in the initial calibration and orientation; software improvements; and the comparison of the MVS models in some studies to reference data captured by time-of-flight laser scanning systems, which might contain as much noise as a photogrammetric cloud.

4.2. Meshed surface and texturing

Next, the software builds a continuous surface mesh over the dense point cloud. The meshes generated from the clouds in Table 5.1 were down-sampled by a factor of 30, resulting in 500,000–800,000 vertices. Though relatively little useful geometric information was lost, errors increased by about 60% (Table 5.2 vs. 5.1). However, a more complex surface, such as an inscription with carved letters, would require higher resolution imagery (Doneus et al. 2011; Koutsoudis et al. 2013; Remondino et al. 2009). The 3-σ error of almost 10 mm was again limited to the crevices, where meshing was unpredictable (Fig. 4b). Otherwise, 95% of the mesh vertices should be plotted accurately within a millimetre. The dimensions of larger features such as overall length or width of a block, which are determined from many thousands of individual point measurements, might be estimated accurately within 0.5 mm or less of the reality.

However, in order to reduce processing times, the dense point clouds for the Heraion were typically rendered below these maximum quality settings. Comparisons were conducted on two pairs of meshes covering long stretches of the cella walls and the outside corners of the stylobate, both rendered from down-sampled imagery. Compared to the full quality models, the error was roughly doubled in the 20-m stretches of the cella walls (Table 5.3), and higher still in the peristyle (Table 5.4) due to large regions that were difficult to photograph. Undercut faces could not

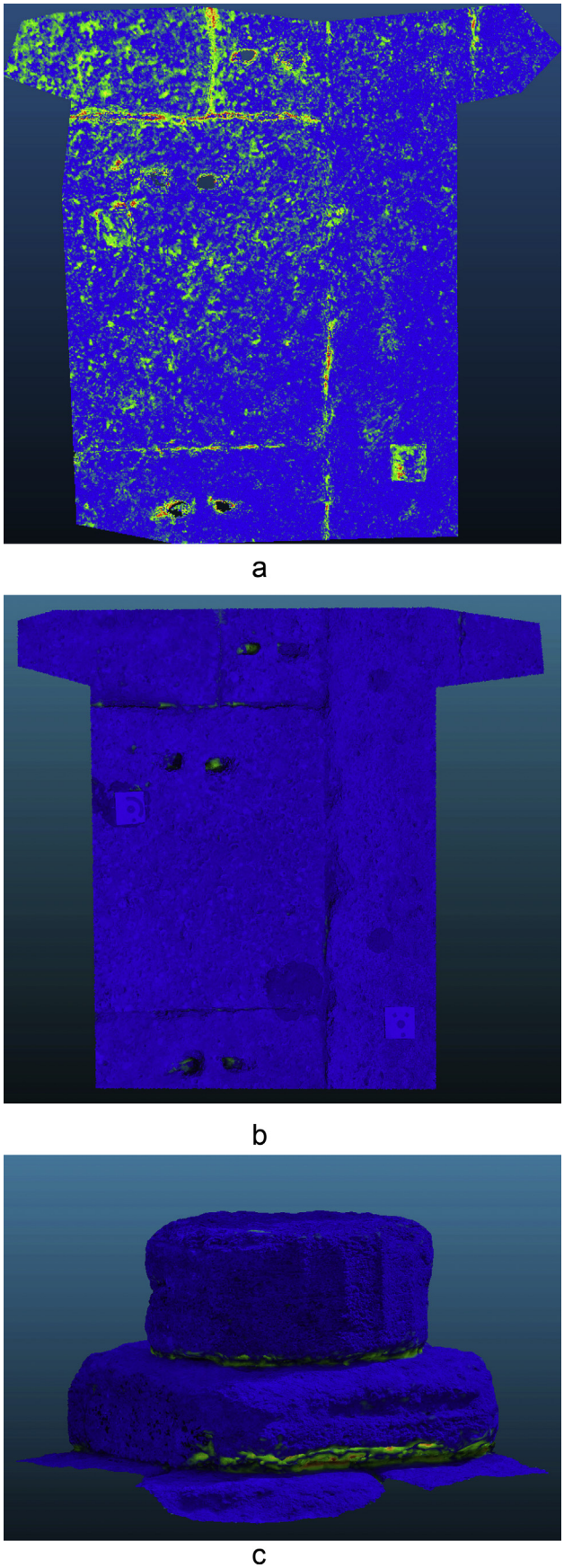

Fig. 4. Dense point cloud and mesh discrepancies (dark/blue = lower errors; light/green-to-red = higher errors). a: Point cloud comparison, NW cella. b: Full quality mesh comparison, NW cella. c: Comparison of 1/16th quality meshes, SW stylobate. Credit: author. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

be accurately captured because they were shaded and too near the ground (Fig. 4c).

Should a higher accuracy be desired in the future, the point clouds can always be reprocessed at higher sampling ratios, but the "Medium" quality meshes met the requirements for the current research. Individual points should be located within 2 mm of the reality at the 2-σ CL, while the dimensions of larger features, like block lengths, can be estimated reliably down to the millimetre. The local precision should be as high as 1:3,500, equivalent to about 2 pixels in the camera sensor.

4.3. Texture accuracy

Finally, the original photographs are projected back onto the modelled surfaces as textures. Any changes in the images (such as an insect straying into the scene) may lead to inaccuracies. There are different approaches to compositing the many overlapping source photographs into a single texture, such as mosaicking or promoting the lightest or darkest pixel values at every position of the model. The preferred method here was averaging, which blends the colour values from all photographs of each area, at a resolution of 1 mm per pixel. Averaging helps control quality. The averaged textures normally show details crisply down to the pixel, but misalignments or bad 3D geometry generate blurring in the textures. Misalignments greater than 1 mm lead to doubled patterns in the surface textures.

The texture accuracy was tested with the same overlapping models at the northwest cella wall. Nine points in the surfaces with distinctive textures were marked manually in 2D on orthographic projections of the texture created from three unrelated sets of photographs taken at different times of day (Fig. 5). Though only 18 points could be compared, they corresponded with an RMS and 2-σ CL of roughly 0.4 and 0.6 mm, respectively. This suggests the 2D texture projections are slightly more internally consistent than the 3D coordinates, which are more difficult to reconstruct than flat 2D projections.

Though conditions over most of the building were ideal for high quality texture reconstruction, less accessible parts of the temple had obvious defects. Blurred textures were common in the lower break faces near the ground, and also the upper parts of the re-erected columns (Fig. 6b vs. a). Their capitals were difficult to photograph and lacked targets. Textures could be mapped incorrectly by a pixel or more, reflecting degraded quality in the models and error in the 1–3 mm range (Fig. 6c–d). New photography on site would be required to reduce these localized errors.

5. Conclusions

As stated in Green et al. 2014, we must assess the accuracy of our photogrammetric models if they are to be treated seriously as a tool for recording and measurement. The quantification of SfM error by discrepancies from Total Station data, compiled in Table 1a, has been shown here to be ineffective. Repeatability testing, though imperfect, is a better method for assessing accuracy and does not require expensive new hardware at the site. Photogrammetry is a complex process, and error will vary in every implementation, making it all the more important for archaeologists to assess general error at the global and local levels.

Any archaeological survey should use coded targets whenever possible. They significantly increase accuracy, reduce the time

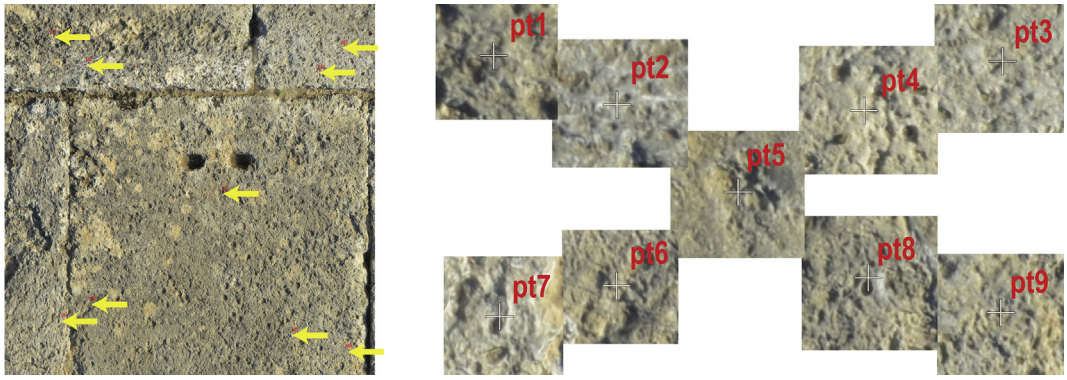

Fig. 5. Texture mapping, details showing the tests points (texture resolution 0.1 mm). Credit: author / Digital Architecture Project.

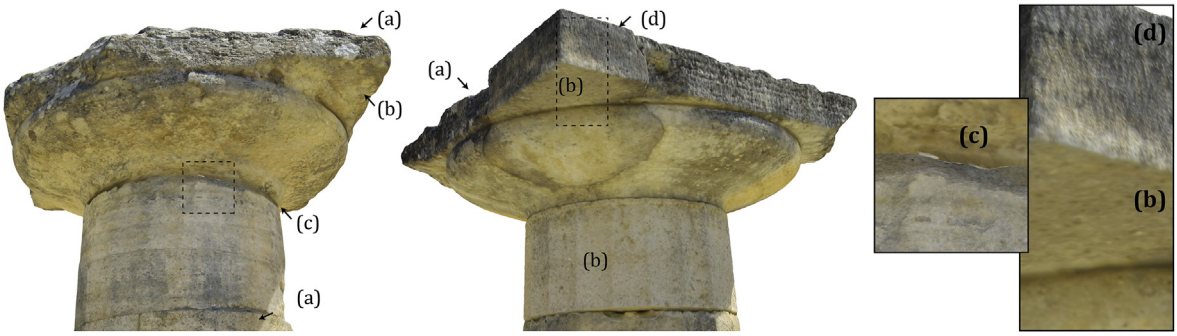

Fig. 6. Problematic texturing at the tops of the columns S15 (left) and E4 (center), with details (right). (a) Areas with adequate imagery have crisp details to the mm; whereas (b) other areas were photographed from the ground at low resolution. Crevices (c) in these areas were incorrectly modelled, and some positions (d) have doubled textures from misalignments of the aerial photographs greater than 1 mm. Credit: author / Digital Architecture Project.

needed to set up jobs in the software, and give us a method to assess error. As shown in Tables 2 and 3, precision is less related to the camera hardware than might be expected. A fixed lens is preferable, but in fact measurement to at least 1:20,000 is possible with a variable lens so long as a significant number of targets is present to assist with the individual calibration for each photograph. With targets, a camera with a full-frame sensor and fixed lens can reach precisions of at least 1:50,000.

The local accuracy of the textured 3D surfaces is not so high. At full quality points can be accurate to as much as 1:8,000, although pitted surfaces will bring the 2-σ CL closer to 1:2000 (Table 5). However, this applies to features small enough to be measured within a single close-up photograph. With the positions of 95% of the coded targets established within 1 mm of the reality during the global survey of the site at Olympia, the clearly visible surface points are accurately located throughout the building within 1.5 mm. Viewed this way, the site-wide accuracy of the models might exceed 1:30,000 at the 2-σ CL, due to the relatively precise initial survey. However, with lower quality settings to save on processing times, most of the models at Olympia have twice this error (95% within 3 mm/1:16,000), which is still an impressive result.

The models from Olympia are significantly more accurate than what could be expected from a mid-range laser scanner, yet they were produced with a camera kit and software that is a small fraction of the cost. Furthermore, the photogrammetric 3D models have photographic texturing showing details down to the millimetre, creating much sharper imagery than laser scanning. By incorporating machine-readable targets into the survey, SfM/MVS is capable of outperforming all other recording methods by cost, detail, and accuracy. These methods thus represent a step toward establishing best practices for high quality 3D recording with photogrammetry.

Acknowledgements

The project at Olympia would have been impossible without the collaboration of Dr. D. Scahill, who supplied and operated the Leica TS for survey on site. M. Creech and O. Hayden assisted during our survey. The work at Olympia was conducted with permission of the Deutsches Archäologisches Institut (DAI) and the 7th Ephorate of Prehistoric and Classical Antiquities at Olympia. Special thanks are due to Dr. R. Senff (DAI) and Dr. G. Chatzi-Spiliopoulou (Ephorate) for their support, including providing accommodations at Olympia, tools, and workmen for cleaning the temple. The Digital Architecture Project is funded by the Department of Art & Art History, the Hixson-Lied College of Fine and Performing Arts, and the Center for Digital Research in the Humanities at the University of Nebraska–Lincoln. The Alpa metric camera whose performance is described in Section 3.1 was loaned to the university by the Atlanta office of Capture Integration, a retailer of medium-format digital cameras.

References

Agisoft PhotoScan Pro: http://www.agisoft.com/pdf/photoscan-pro_1_1_en.pdf.
Barazzetti, L., Mussio, L., Remondino, F., Scaioni, M., 2011. Targetless camera calibration. In: Remondino, F., El-Hakim, S. (Eds.), 4th ISPRS Int. Workshop 3D-ARCH 2011 (Vol. XXXVIII-5/W16), pp. 335–342. http://dx.doi.org/10.5194/isprsarchives-XXXVIII-5-W16-335-2011.
CloudCompare: http://www.danielgm.net/cc/.
Dai, F., Lu, M., 2010. Assessing the accuracy of applying photogrammetry to take

geometric measurements on building products. J. Constr. Eng. Manag. 136, 242–250. http://dx.doi.org/10.1061/(ASCE. http://dx.doi.org/10.1061/(ASCE)CO.1943-7862.0000114.

De Reu, J., De Smedt, P., Herremans, D., Van Meirvenne, M., Laloo, P., De Clercq, W., 2014. On introducing an image-based 3D reconstruction method in archaeological excavation practice. J. Archaeol. Sci. 41, 251–262. http://dx.doi.org/10.1016/j.jas.2013.08.020.

De Reu, J., Plets, G., Verhoeven, G., De Smedt, P., Bats, M., Cherretté, B., De Maeyer, W., Deconynck, J., Herremans, D., Laloo, P., Van Meirvenne, M., De Clercq, W., 2013. Towards a three-dimensional cost-effective registration of the archaeological heritage. J. Archaeol. Sci. 40, 1108–1121. http://dx.doi.org/10.1016/j.jas.2012.08.040.

Dellepiane, M., Dell'Unto, N., Callieri, M., Lindgren, S., Scopigno, R., 2013. Archeological excavation monitoring using dense stereo matching techniques. J. Cult. Herit. 14, 201–210. http://dx.doi.org/10.1016/j.culher.2012.01.011.

Doneus, M., Verhoeven, G., Fera, M., Briese, Ch, Kucera, M., Neubauer, W., 2011. From deposit to point cloud – a study of low-cost computer vision approaches for the straightforward documentation of archaeological excavations. In: Čepek, A. (Ed.), Geoinformatics. XXIIIrd Int. CIPA Symp, vol. 6. Czech University, Prague, pp. 81–88. http://dx.doi.org/10.14311/gi.6.11.

Dörpfeld, W., Schleif, H., 1935. Alt-Olympia. Untersuchungen und Ausgrabungen zur Geschichte des ältesten Heiligtums von Olympia und der älteren griechischen Kunst. E.S. Mittler & Sohn, Berlin.

Fernández-Hernandez, J., González-Aguilera, D., Rodríguez-González, P., Mancera-Taboada, J., 2014. Image-based modelling from unmanned aerial vehicle (UAV) photogrammetry: an effective, low-cost tool for archaeological applications. Archaeom 57 (1), 128–145. http://dx.doi.org/10.1111/arcm.12078.

Fraser, C.S., 1992. Photogrammetric measurement to one part in a million. Photogramm. Eng. Remote Sens. 58, 305–310.

Fraser, C.S., 2012. Automatic camera calibration in close-range photogrammetry. In: ASPRS 2012 Annual Conf. Sacramento, CA March 19–23, 2012. In: http://info.asprs.org/publications/proceedings/Sacramento2012/files/Fraser.pdf.

Fraser, C.S., Al-Ajlouni, S., 2006. Zoom-dependent camera calibration in close-range photogrammetry. Photogramm. Eng. Remote Sens. 72 (9), 1017–1026.

Fraser, C.S., Brown, D.C., 1986. Industrial photogrammetry: new developments and recent applications. Photogramm. Rec. 12 (68), 197–217. http://dx.doi.org/10.1111/j.1477-9730.1986.tb00557.x.

Fraser, C.S., Woods, A., Brizzi, D., 2005. Hyper redundancy for accuracy enhancement in automated close range photogrammetry. Photogramm. Rec. 20, 205–217. http://dx.doi.org/10.1111/j.1477-9730.2005.00327.x.

Galantucci, L.M., Lavecchia, F., Percoco, G., Raspatelli, S., 2014. New method to calibrate and validate a high-resolution 3D scanner, based on photogrammetry. Precis. Eng. 38 (2), 279–291. http://dx.doi.org/10.1016/j.precisioneng.2013.10.002.

Green, S., Bevan, A., Shapland, M., 2014. A comparative assessment of structure from motion methods for archaeological research. J. Archaeol. Sci. 46, 173–181. http://dx.doi.org/10.1016/j.jas.2014.02.030.

Höhle, J., Höhle, M., 2009. Accuracy assessment of digital elevation models by means of robust statistical methods. ISPRS J. Photogramm. Remote Sens. 64, 398–406. http://dx.doi.org/10.1016/j.isprsjprs.2009.02.003.

Jennings, A., Black, J., 2012. Texture-based photogrammetry accuracy on curved surfaces. AIAA J. 50 (5), 1060–1071. http://dx.doi.org/10.2514/1.J050956.

Kersten, T.P., Lindstaedt, M., 2012. Image-based low-cost systems for automatic 3D recording and modelling of archaeological finds and objects. In: Ioannides, M., Fritsch, D., Leissner, J., Davies, R., Remondino, F., Caffo, R. (Eds.), Progress in Cultural Heritage Preservation. EuroMed 2012. Springer, Berlin, pp. 1–10. http://dx.doi.org/10.1007/978-3-642-34234-9_1.

Koutsoudis, A., Vidmar, B., Arnaoutoglou, F., 2013. Performance evaluation of a multi-image 3D reconstruction software on a low-feature artefact. J. Archaeol. Sci. 40, 4450–4456. http://dx.doi.org/10.1016/j.jas.2013.07.007.

Koutsoudis, A., Vidmar, B., Ioannakis, G., Arnaoutoglou, F., Pavlidis, G., Chamzas, C., 2014. Multi-image 3D reconstruction data evaluation. J. Cult. Herit. 15, 73–79. http://dx.doi.org/10.1016/j.culher.2012.12.003.

Lerma, J.L., Muir, C., 2014. Evaluating the 3D documentation of an early Christian upright stone with carvings from Scotland with multiple images. J. Archaeol. Sci. 46, 311–318. http://dx.doi.org/10.1016/j.jas.2014.02.026.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60 (2), 91–110. http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94.

Luhmann, T., 2010. Close range photogrammetry for industrial applications. ISPRS J. Photogramm. Remote Sens. 65, 558–569. http://dx.doi.org/10.1016/j.isprsjprs.2010.06.003.

Martínez, S., Ortiz, J., Gil, M., 2015. Geometric documentation of historical pavements using automated digital photogrammetry and high-density reconstruction algorithms. J. Archaeol. Sci. 53, 1–11. http://dx.doi.org/10.1016/j.jas.2014.10.003.

Olson, B.R., Placchetti, R.A., Quartermaine, J., Killebrew, A.E., 2013. The Tel Akko total archaeology project (Akko, Israel): assessing the suitability of multi-scale 3D field recording in archaeology. J. Field Archaeol. 38 (3), 244–262. http://dx.doi.org/10.1179/0093469013Z.00000000056.

Reinoso, J.F., Moncayo, M., Barrera, D., 2014. Close-range photogrammetry applied to the documentation of cultural heritage using telescopic and wide-angle lenses. Imaging Sci. J. 62, 387–394. http://dx.doi.org/10.1179/1743131X14Y.0000000077.

Remondino, F., del Pizzo, S., Kersten, T.P., Troisi, S., 2012. Low-cost and open-source solutions for automated image orientation – a critical overview. In: Ioannides, M., Fritsch, D., Leissner, J., Davies, R., Remondino, F., Caffo, R. (Eds.), Progress in Cultural Heritage Preservation. EuroMed 2012. Springer, Berlin, pp. 40–54. http://dx.doi.org/10.1007/978-3-642-34234-9_5.

Remondino, F., El-Hakim, S., Baltsavias, E., Picard, M., Grammatikopoulos, L., 2008. Image-based 3D modeling of the Erechtheion, Acropolis of Athens. In: Chen, J., Jiang, J., Maas, H.G. (Eds.), XXIst ISPRS Cong. Tech. Commission V (Vol. XXXVII-B5), pp. 1083–1092.

Remondino, F., Girardi, S., Rizzi, A., Gonzo, L., 2009. 3D modeling of complex and detailed cultural heritage using multi-resolution data. J. Comput. Cult. Herit. 2.1 (2), 1–20. http://dx.doi.org/10.1145/1551676.1551678.

Rieke-Zapp, D., Tecklenburg, W., Peipe, J., Hastedt, H., Haig, C., 2009. Evaluation of the geometric stability and the accuracy potential of digital cameras — comparing mechanical stabilisation versus parameterisation. ISPRS J. Photogramm. Remote Sens. 64 (3), 248–258. http://dx.doi.org/10.1016/j.isprsjprs.2008.09.010.

Riveiro, B., Caamaño, J.C., Arias, P., Sanz, E., 2011. Photogrammetric 3D modelling and mechanical analysis of masonry arches: an approach based on a discontinuous model of voussoirs. Autom. Constr. 20, 380–388. http://dx.doi.org/10.1016/j.autcon.2010.11.008.

Roosevelt, C.H., Cobb, P., Moss, E., Olson, B.R., Ünlüsoy, S., 2015. Excavation is ~~destruction~~ digitization: advances in archaeological practice. J. Field Archaeol. 40 (3), 325–346. http://dx.doi.org/10.1179/2042458215Y.0000000004.

Sapirstein, P., 2015. Photogrammetry as a tool for architectural analysis: the digital architecture project at Olympia. In: Papadopoulos, C., et al. (Eds.), Archaeological Research in the Digital Age. Proceedings of the 1st Conference on Computer Applications and Quantitative Methods in Archaeology, Greek Chapter. Rethymno, Crete, March 6–8 2014, pp. 129–139.

Stal, C., Van Liefferinge, K., De Reu, J., Docter, R., Dierkens, G., De Maeyer, P., Mortier, S., Nuttens, T., Pieters, T., van den Eijnde, F., van de Put, W., De Wulf, A., 2014. Integrating geomatics in archaeological research at the site of Thorikos. J. Archaeol. Sci. 45, 112–125. http://dx.doi.org/10.1016/j.jas.2014.02.018.

Stamatopoulos, C., Fraser, C.S., 2011. Calibration of long focal length cameras in close range photogrammetry. Photogramm. Rec. 26, 339–360. http://dx.doi.org/10.1111/j.1477-9730.2011.00648.x.

Stamatopoulos, C., Fraser, C.S., 2014. Automated target-free network orientation and camera calibration. In: Remondino, F., Menna, F. (Eds.), ISPRS Tech. Commission V. Symp. (Vol. XL-5), pp. 339–346. http://doi.org/doi:10.5194/isprsannals-II-5-339-2014.

Toschi, I., Capra, A., De Luca, L., Beraldin, J.A., Cournoyer, L., 2014. On the evaluation of photogrammetric methods for dense 3D surface reconstruction in a metrological context. In: Remondino, F., Menna, F. (Eds.), ISPRS Tech. Commission V. Symp. (Vol. XL-5), pp. 371–378. http://doi.org/10.5194/isprsannals-II-5-371-2014.

Vergauwen, M., Van Gool, L., 2006. Web-based 3D reconstruction service. Mach. Vis. Appl. 17, 411–426. http://dx.doi.org/10.1007/s00138-006-0027-1.

Zhenzhong, X., Liang, J., Yu, D., Tang, Z., Asundi, A., 2010. An accurate stereo vision system using cross-shaped target self-calibration method based on photogrammetry. Opt. Lasers Eng. 48, 1252–1261. http://dx.doi.org/10.1016/j.optlaseng.2010.06.006.