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 xxx

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.