Journal of Archaeological Science 49 (2014) 249–254

Contents lists available at ScienceDirect

Journal of Archaeological Science

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

Focus article

Computer vision, archaeological classification and China's terracotta warriors

Andrew Bevan [a, *], Xiuzhen Li [a, b], Marcos Martinón-Torres [a], Susan Green [a], Yin Xia [b], Kun Zhao [b], Zhen Zhao [b], Shengtao Ma [b], Wei Cao [b], Thilo Rehren [a, c]

[a] Institute of Archaeology, University College London, London, UK
[b] Emperor Qin Shihuang's Mausoleum Site Museum, Xi'an, China
[c] UCL-Q, Doha, Qatar

ARTICLE INFO

Article history:
Received 18 January 2014
Received in revised form
14 May 2014
Accepted 19 May 2014
Available online xxx

Keywords:
Geomorphometrics
Biometrics
Photogrammetry
Craft production

ABSTRACT

Structure-from-motion and multiview-stereo together offer a computer vision technique for reconstructing detailed 3D models from overlapping images of anything from large landscapes to microscopic features. Because such models can be generated from ordinary photographs taken with standard cameras in ordinary lighting conditions, these techniques are revolutionising digital recording and analysis in archaeology and related subjects such as palaeontology, museum studies and art history. However, most published treatments so far have focused merely on this technique's ability to produce low-cost, high quality representations, with one or two also suggesting new opportunities for citizen science. However, perhaps the major artefact scale advantage comes from significantly enhanced possibilities for 3D morphometric analysis and comparative taxonomy. We wish to stimulate further discussion of this new research domain by considering a case study using a famous and contentious set of archaeological objects: the terracotta warriors of China's first emperor.

1. Introduction

Structure-from-motion and multiview-stereo (SfM–MVS) together constitute a computer vision approach to creating 3D colour-realistic models from a series of overlapping digital photographs (Szeliski, 2011). In archaeology, SfM–MVS are revolutionising the nature of recording and analysis of archaeological artefacts, sites and landscapes (Ducke et al., 2011; Remondino et al., 2012; Verhoeven et al, 2012; Olson et al., 2013), with similar reverberations in related subjects such as paleontology, art history and museum studies. However, most treatments so far have emphasized high fidelity primary documentation, some preliminary considerations of model accuracy or preferred software, and some opportunities for 'citizen science' (Snaveley et al., 2008). In addition, we would stress a further key application that has received little or no archaeological attention to date, but which will have particularly important implications for a core archaeological endeavour: the classification of artefacts. We consider this opportunity below with reference to the Qin terracotta warriors, perhaps the most well-known representatives of China's most famous archaeological site, the mausoleum of the first Chinese emperor, Qin Shihuangdi (259–210 BC; SIAATQ, 1988; Yuan, 1990; Portal, 2007). Our preliminary study below draws on a selection of warriors from the most extensively investigated part of the complex and the most widely known group of terracotta warriors, Pit 1, and is part of an ongoing cooperative project studying the construction methods and logistical organisation underpinning the terracotta army and Qin Shihuang mausoleum complex, especially from the perspective of materials science, shape analysis and spatial modelling (e.g. Li et al., 2011; Martinón-Torres et al., 2013; Bevan et al., 2013; Li et al., 2014).

2. Model construction

The 3D models of warriors that are considered here reflect the best results achieved via both open source and proprietary software implementations of SfM–MVS (VisualSFM and Photoscan, as well as Meshlab, CloudCompare and R for further processing or analysis), using a range of parameter choices. SfM–MVS software can be used on a consumer grade laptop or ordinary desktop computer, but it does make heavy computational demands. For example, on a 64-bit computer with 64 GB of RAM, a 1 GB GPU and a six-core

* Corresponding author.
E-mail address: a.bevan@ucl.ac.uk (A. Bevan).

http://dx.doi.org/10.1016/j.jas.2014.05.014

3.20 GHz CPU, a head-and-shoulders model of a warrior from ca.25 photographs takes a few minutes to complete while a model of a full warrior from ca.100 photographs takes several hours, excluding model clean-up and simplification.

A typical SfM–MVS process involves several steps: image creation or collection, feature detection and matching, sparse bundle reconstruction, and thereafter optionally, dense point cloud reconstruction, mesh construction and photo-texturing. These steps have already received some attention from archaeologists elsewhere so we will only briefly summarise them here. Ordinary photographs provide the initial input data for SfM–MVS models and these can either be acquired from existing collections or captured fresh. For the individual warrior and warrior ear models, we collected a new set of photographs taken with a modern digital SLR (without tripod) under normal daytime lighting conditions in Pit 1. Significant overlap between images is a key prerequisite for success (Fig. 1a) and we collected horizontal bands of images at approximately 15° offsets (i.e. 24 per band) and with further vertical overlap between bands. After image acquisition, the (fully-automatic) SfM–MVS process begins by assessing each photograph to identify distinct groups of pixels that constitute features that are likely to be discernible in several images (Lowe, 2004). After these features have been described for each image, they are matched across multiple images to produce a network of spatial relationships from which individual camera position for each photograph can be reconstructed. The end result is a sparse cloud of 3D point locations that mark the successfully matched features (Fig. 1b). Thereafter, a much denser set of 3D points can be created by grouping the image sequence into sub-sequences of images covering similar parts of the surface and then looking for more detailed feature matches over a coarse search grid (Furukawa and Ponce, 2010). Parameter choices such as the minimum necessary number of matched features or the size of the dense search grid affect the resulting number and quality of reconstructed points, as well as the overall computational requirements. The 3D point clouds generated via the above steps also contain the colour information from the original image pixels, as well as a degree of noise that might be due to unwanted additional objects in the photos, occasional atmospheric effects or variegated background. Such rogue features can be deleted or masked prior to matching and/or removed manually afterwards. An SfM approach does not begin with any inherent sense of the spatial scale or geographic location of the (otherwise geometrically accurate) model it creates, and this needs to be added in a further step, either by marking points on the photographs prior to model construction or re-scaling and georeferencing the model afterwards. If required, a triangular mesh version can also be made via several alternative methods (e.g. Kazhdan and Hoppe, 2013) and detailed photographic texture can be applied per face instead of averaged colour.

Traditionally, archaeologists have recorded sites and artefacts via a combination of ordinary still photographs, 2D line drawings and occasional cross-sections. Given these constraints, the attractions of 3D models have been obvious for some time, with digital photogrammetry and laser scanners offering two well-known methods for data capture at close range (e.g. Bates et al., 2010; Hess and Robson, 2010). The highest specification laser scanners still boast better positional accuracy and greater true colour fidelity than SfM–MVS methods (James and Robson, 2012), but the latter

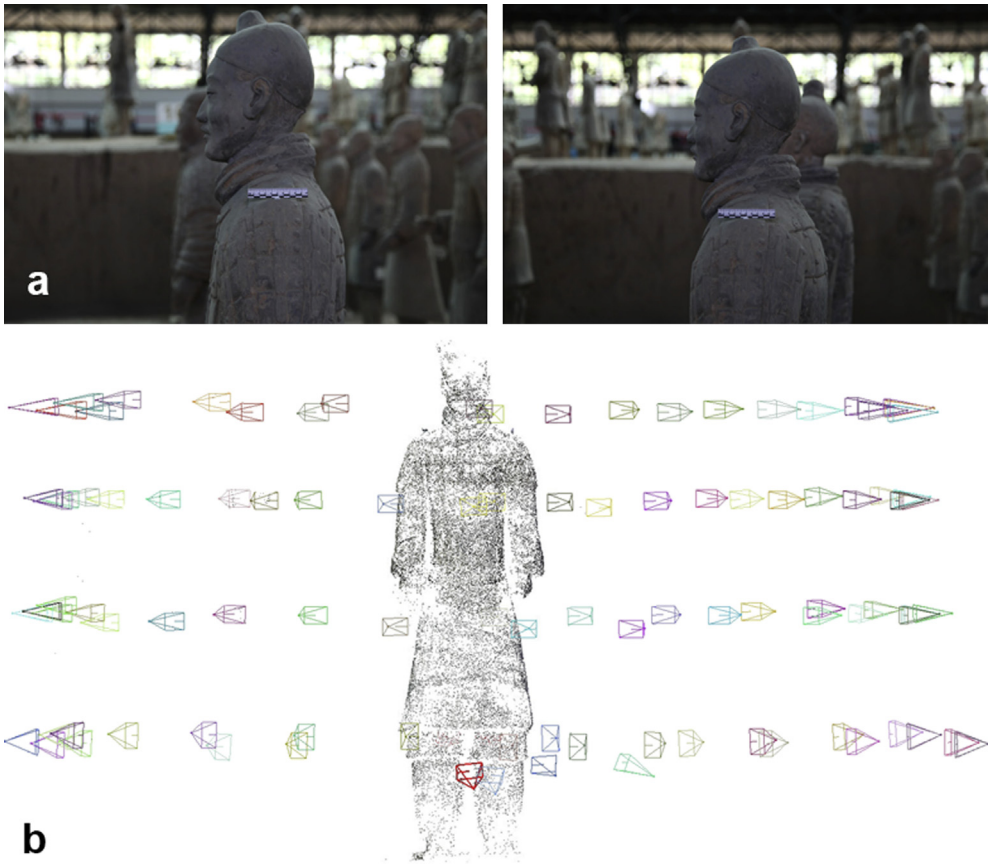

Fig. 1. (a) Two example photographs of a warrior taken at slight offsets, out of a larger set covering this entire warrior, and (b) a sparse point cloud of the same warrior along with reconstructed camera positions.

produce very good quality models nonetheless and have many unique selling points. Unlike traditional digital photogrammetry, little or no prior control of camera position is necessary, and unlike laser scanning, no major equipment costs or setup are involved. However, the key attraction of SfM–MVS is that the required input can be taken by anyone with a digital camera and modest prior training about the required number and overlap of photographs. A whole series of traditional bottlenecks are thereby removed from the recording process and large numbers of archaeological landscapes, sites or artefacts can now be captured rapidly, in the field, in the laboratory or in the museum. Fig. 2a–c shows examples of terracotta warrior models for which the level of surface detail is considerable.

3. 3D shape analysis

Beyond high quality visualisation, we would argue that perhaps the most compelling analytical rationale for SfM–MVS is that it can be scaled up to capture, not just one or two artefacts, but large numbers of 3D models whose surface geometries can be formally

Fig. 2. (a) a textured 3D mesh of a warrior, (b–c) close-ups of two other 3D models of warrior's faces.

compared. Such 3D morphometric analysis has simply not been possible in the past, because of the often prohibitive purchase costs, lack of expert operators and difficult set-ups typical for laser-scanning, but SfM–MVS now provides a ready solution. As an example, one way to assess variability in the micro-style and construction techniques of individual terracotta warriors is to consider the shape of features such as faces, hands or ears across a range of warriors. Ear morphology exhibits strong variation amongst real humans to the extent that it has been used to identify individuals and in forensic work for over a century (Bertillon, 1893; Pflug and Busch, 2012; Abaza et al., 2013). Ear biometrics are also of great interest to human geneticists (Hunter et al., 2009). On the other hand, artistic renderings of human ears present a more complex case. A famous early application of scientific method to art history was the Italian art critic Giovanni Morelli's (1892–3) suggestion that incidental details of the way a particular artist portrays ears and hands might be used to attribute unsigned paintings or sculptures to know artists ('Morellian' method; Wollheim, 1973; Ginzburg, 1980). The terracotta warriors' ears were made from the same loess-rich, pale clays as the rest of the figures' bodies and were probably hand-finished at a fairly late stage in their manufacture. Fig. 3 demonstrates that there are visible differences in the way they were rendered on different warriors. This variation could conceivably relate to the signature working habits of particular artisans, to the makers' desire for warriors that exhibited a realistic degree of anatomical individualism or to a situation in which the warriors were actually portraits of real individuals (for discussion of the latter suggestion, Kesner, 1995).

Typically, the statistical analysis of complex shapes such as those exhibited by biological organisms has involved identification of perceived 'landmarks' on the subject (or semi-landmarks anchored to these) and then comparison of such sparse 2D or 3D

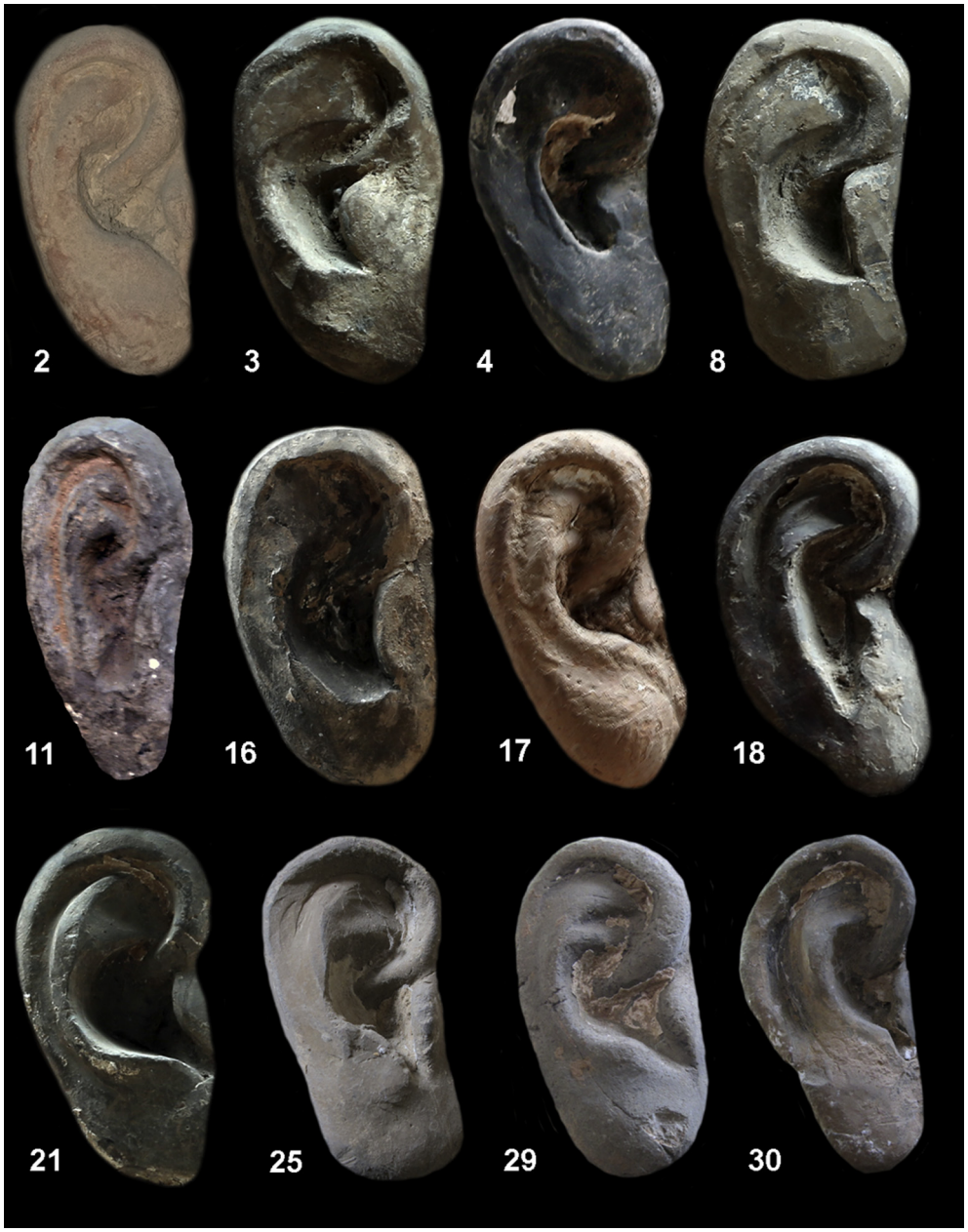

Fig. 3. Examples of twelve different ears (heights have been standardised). The numbers cross-reference with Fig. 5 and the Supplementary dataset.

point sets via Procrustes superimposition (Dryden and Mardia, 1998; Mitteroecker and Gunz, 2009). However, it is not always obvious how any such landmarks could be reliably chosen on the continuous surface geometry of something like an ear. An increasingly popular alternative is to anchor a string of 'semi-landmarks' onto a few real landmarks, either just for 2D outlines (Monna et al., 2013) or indeed for 3D surfaces (MacLeod, 2010). However, even this level of prior knowledge about appropriate anchor points is sometimes problematic and there are also increasing calls to adopt landmark-free methods for complex shapes such as ears, using dense 3D point clouds (Yan and Bowyer, 2007; Wuhrer et al., 2011).

As a preliminary foray with the rather specific demands of archaeological data in mind, we wish to propose a method for constructing a distance matrix that expresses the pairwise dissimilarity of artefacts to others in an assemblage. Distance matrices are common building blocks underpinning well-known statistical clustering and ordination methods, and would also enable phylogenetic analysis in cases where branching evolutionary relationships might be hypothesised. Below, we suggest that shape differences among 3D models of artefacts can be expressed via a matrix built up by calculating the mean or median distance between each point in one cloud and its nearest neighbour in another, once both clouds have been finely co-registered with one another.

As an example to fix ideas, we took photographs of the faces of 30 warriors from one side, deliberately avoiding too many ear close-ups as our ultimate purpose is to document the entire group of 1000+ excavated warriors without moving them from their crowded location down in the mausoleum pits. The resulting point clouds are detailed but not exceptionally so, and any analytical technique for comparing ear surface geometries needs to handle a limited number of small gaps where SfM–MVS was unable to find sufficient feature matches or where actual ear anatomy is obscured by bits of unexcavated soil (e.g. sometimes in parts of the auricular well). We chose to extract each ear's point cloud from a wider model of the warriors face and then to standardise the model's size, position, orientation and point density (Fig. 4a–b). More precisely, we realigned the ear point cloud to the XY plane via a least squares regression (known as an n-point strike-and-dip method in geology: Fienen, 2005), reflected left ears to become right ears, rotated, rescaled and centred each model to both a common unit height and origin. This allows a more straightforward comparison between any two ears, in which they are both of unit height and oriented the same way (very much the same pre-processing used for 2D and outline-based morphometrics). To further ensure fair comparison between models, we also down-sampled each point cloud to a consistent point density.

Once every ear is represented by a standardised point cloud (see the Supplementary data for these) it can be more finely co-registered with every other one in turn (e.g. Fig. 4c), using an iterative closest point (ICP) method (Besl and McKay, 1992). First, one model (*X*) is designated the 'data' and the other (*Y*) the 'target', to which *X* will be finely registered. The ICP process begins by finding a set of points in *Y* which represent the closest neighbours to each point in *X* and, based on this, then computes a least squares transformation of *X* to *Y*, along with an accompanying measure of mean square error. A new set of closest points on Y can then be calculated, and the iteration continues until an agreed threshold of convergence (i.e. until the observed error ceases to change much).

At convergence, this summary mean square statistic or a similar one can be used to express the goodness-of-fit between the two models and as a global measure of pairwise dissimilarity to populate a complete distance matrix. The resulting distances among all ear pairs can be visualised via an ordination method such as multi-

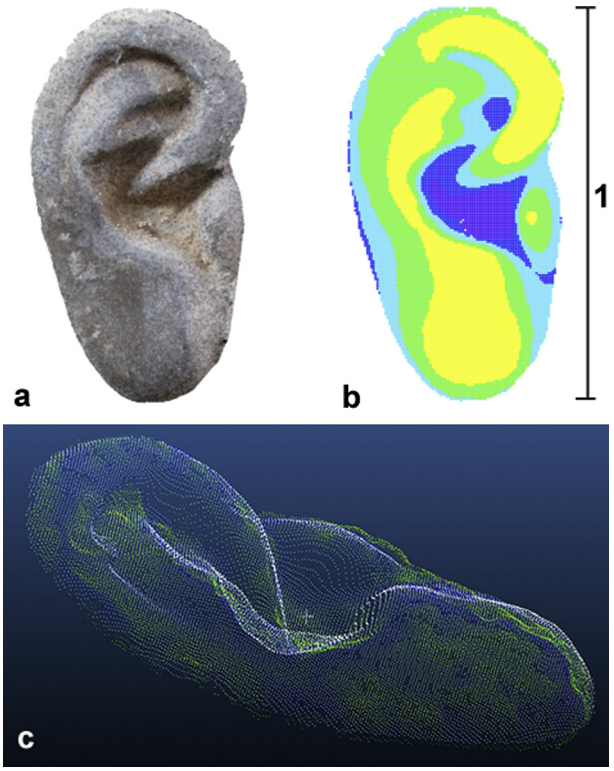

Fig. 4. (a) A dense point cloud of an ear (28), (b) the same ear standardised from comparative purposes by size, orientation and point density (coloured for relative height above the plane), and (c) fine co-registration of this ear (shown in blue–green) to another ear model (shown in white). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

dimensional scaling (Fig. 5), but would also further support hierarchical clustering or phylogenetic modelling. The approach will not necessarily produce symmetric results so needs to be calculated in both directions for each artefact pair (i.e. switching which

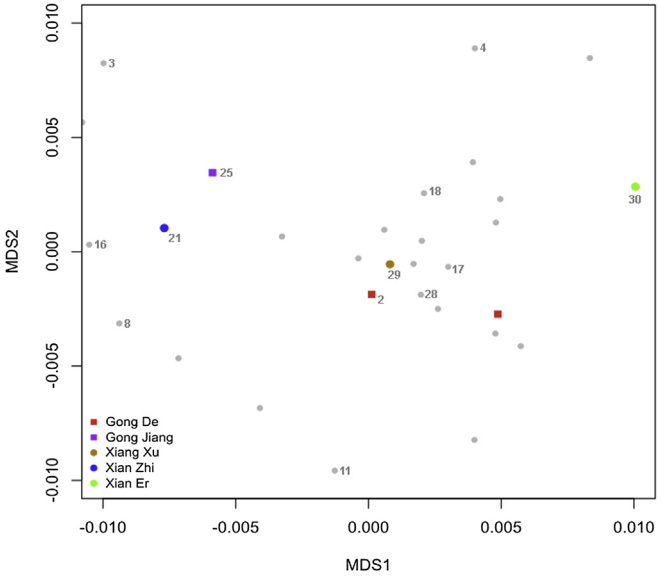

Fig. 5. A multi-dimensional scaling of ear dissimilarities. The coloured symbols are those warriors in the sample that have one or more inscriptions on their surface (as shown in the legend, with the location first, then the foreman's name). Numbered symbols cross-reference with Fig. 3 and the Supplementary dataset. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

artefact is *X* and *Y*). In principle the same method can be used for other kinds of 3D model (e.g. triangular meshes or 'solid' boundary representations) if they can be decomposed into or approximated by a point cloud. One extension of this technique would be to allow localized weighting of the points, such that parts of the artifact can be fitted separately and an overall mapping of areas of better and worse fit across the artefact can be made.

4. Discussion

Our initial results with this sample of ears from the Qin terracotta warriors strongly indicates that, while there is a core of approximately similar shapes (e.g. Fig. 3 (2, 17, 29), Fig. 4a (28)), there is also considerable variation to the extent that, in contrast to the warriors' highly standardised bronze weapons for example (Martinón-Torres et al., 2013; Li et al., 2014), no two ears are strictly the same. Likewise, there is as yet little evidence for a close relationship between ear microstyle and the limited inscriptional evidence naming the foreman in charge of the construction of a particular warrior or the occasional mention of place-names (perhaps distinct workshop locales or worker origins) such as the capital at Xianyang ("Xian") or the imperial palace ("Gong"). This tentatively supports the hypothesis that the warriors were intended to constitute a real army, whose weapons were standardised (and lethal) but whose individual soldiers were not. It remains to be seen whether the ears themselves exhibit comparable levels of individuality to what we might expect in a real population of adult males (as seems likely from the warrior height distribution: Komlos, 2003) and/or whether subtle indications of workshop microstyle or spatial clusters in the pit are visible in larger samples. At any rate, it should be clear that, beyond low-cost high quality documentation and novel citizen science, SfM–MVS also enables more flexible approaches to 3D shape analysis that have the potential to revolutionise artefact classification and scientific taxonomy in archaeology over the next few years.

Acknowledgements

This research was conducted as part of the cooperative *Imperial Logistics* project between the Emperor Qin Shihuang's Mausoleum Site Museum and the Institute of Archaeology, University College London. Amongst many contributors beyond those listed as authors above, we are particularly indebted to the late Peter Ucko, Stephen Shennan and Wu Yongqi for facilitating the smooth development of this collaboration. We are also grateful to the British Academy for financial and institutional support in making this one of its official Research Projects. Xiuzhen Li's contribution was also supported by a grant from Rio Tinto PLC. Thanks also to Benjamin Ducke, Daniel Girardeau-Montaut and Changchang Wu who offered helpful suggestions.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jas.2014.05.014.

References

Abaza, A., Ross, A., Hebert, C., Harrison, M.A.F., Nixon, M.S., 2013. A survey on ear biometrics. ACM Comput. Surv. 45, 22.
Bates, K.T., Falkingham, P.L., Rarity, F., Hodgetts, D., Purslow, A., Manning, P.L., 2010. Application of high-resolution laser scanning and photogrammetric techniques to data acquisition, analysis and interpretation in palaeontology. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 38, 68–73.
Bertillon, A., 1893. Instructions Signalétiques. Melun: Imprimerie Administrative.
Besl, P.J., McKay, N.D., 1992. A method for registration of 3D shapes. IEEE Trans. Pattern Anal. Mach. Intell. 14, 239–256.
Bevan, A., Crema, E., Li, X.J., Palmisano, A., 2013. Intensities, interactions and uncertainties: some new approaches to archaeological distributions. In: Bevan, A., Lake, M. (Eds.), Computational Approaches to Archaeological Spaces. Left Coast Press, Walnut Creek, pp. 27–51.
Dryden, I.L., Mardia, K.V., 1998. Statistical Shape Analysis. Wiley, New York.
Ducke, B., Score, D., Reeves, J., 2011. Multiview 3D reconstruction of the archaeological site at Weymouth from image series. Comput. Graph. 35, 375–382.
Fienen, M.N., 2005. The three-point problem, vector analysis and extension to the n-point problem. J. Geosci. Educ. 53 (3), 257–262.
Furukawa, Y., Ponce, J., 2010. Accurate, dense, and robust multi-view stereopsis. IEEE Trans. Pattern Anal. Mach. Intell. 32, 1362–1376.
Ginzburg, C., 1980. Morelli, Freud and Sherlock Holmes: clues and scientific method. Hist. Workshop 9, 5–36.
Hess, M., Robson, S., 2010. 3D colour imaging for cultural heritage artefacts. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 38 (5), 288–292.
Hunter, A., Frias, J.L., Gillessen-Kaesbach, G., Hughes, H., Lyons Jones, K., Wilson, L., 2009. Elements of morphology: standard terminology for the ear. Am. J. Med. Genet. A 149A, 40–60.
James, M.R., Robson, S., 2012. Straightforward reconstruction of 3D surfaces and topography with a camera: accuracy and geoscience application. J. Geophys. Res. 117, F03017.
Kazhdan, M., Hoppe, H., 2013. Screened Poisson surface reconstruction. ACM Trans. Graph. 32, e29.
Kesner, L., 1995. Likeness of No One: (re)presenting the First Emperor's army. Art. Bull. 77, 115–132.
Komlos, J., 2003. The size of the Chinese terracotta warriors – 3rd century B.C. Antiquity 77. Available: http://antiquity.ac.uk/projgall/komlos/komlos.html (accessed 02.10.13).
Li, X.J., Martinón-Torres, M., Meeks, N.D., Xia, Y., Zhao, K., 2011. Inscriptions, filing, grinding and polishing marks on the bronze weapons from the Qin Terracotta Army in China. J. Archaeol. Sci. 38, 492–501.
Li, X.J., Bevan, A., Martinón-Torres, M., Rehren, T., Cao, W., Xia, Y., Zhao, K., 2014. Crossbows and imperial craft organisation: the bronze triggers of China's Terracotta Army. Antiquity 88 (339), 126–140.
Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60, 91–110.
MacLeod, N., 2010. Alternative 2D and 3D form characterization approaches to the automated identification of biological species. In: Nimis, P.L., Vignes Lebbe, R. (Eds.), Tools for Identifying Biodiversity: Progress and Problems. University of Trieste, Trieste, pp. 225–229.
Martinón-Torres, M., Li, X.J., Bevan, A., Xia, Y., Zhao, K., Rehren, T., 2013. Forty thousand arms for a single emperor: from chemical data to the labour organization behind the bronze arrows of the Terracotta Army. J. Archaeol. Method Theory (online).
Mitteroecker, P., Gunz, P., 2009. Advances in geometric morphometrics. Evol. Biol. 36, 235–247.
Monna, F., Jebrane, A., Gabillot, M., Laffont, R., Specht, M., Bohard, B., Camizuli, E., Petit, C., Chateau, C., Alibert, P., 2013. Morphometry of Middle Bronze Age palstaves. Part II- spatial distribution of shapes in two typological groups, implications for production and exportation. J. Archaeol. Sci. 40, 507–516.
Morelli, G., 1892–3. Italian Painters: Critical Studies of their Works. John Murray, London.
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.
Pflug, A., Busch, C., 2012. Ear biometrics: a survey of detection, feature extraction and recognition methods. IET Biom. 1, 114–129.
Portal, J. (Ed.), 2007. The First Emperor: China's Terracotta Army. British Museum, London.
Remondino, et al., 2012. Low-cost and open-source solutions for automated image orientation – a critical overview. In: Progress in Cultural Heritage Preservation. Proceedings of the 4th International Conference, EuroMed 2012. Springer, Berlin, Heidelberg, pp. 40–54. Euromed2012, Lemessos, Cyprus. October 29-November 3, 2012.
Snavely, N., Seitz, S.M., Szeliski, R., 2008. Modeling the world from Internet photo collections. Int. J. Comput. Vis. 80, 189–210.
Szeliski, R., 2011. Computer Vision. Algorithms and Applications. Springer, New York.
SIAATQ – Shaanxi Institute of Archaeology and Archaeological Team of Qinshihunagling, 1988. Qin Shihuang Ling Bingmayong Keng – Yihao Keng Fajue Baogao 1974–1984. Cultural Relics Press, Beijing.
Verhoeven, G., Doneus, M., Briesec, C., Vermeulen, F., 2012. Mapping by matching: a computer vision-based approach to fast and accurate georeferencing of archaeological aerial photographs. J. Archaeol. Sci. 39, 2060–2070.
Wollheim, R., 1973. Giovanni Morelli and the origins of scientific connoisseurship. In: Wollheim, R. (Ed.), On Art and the Mind: Essays and Lectures. Allen Lane, London, pp. 177–201.
Wuhrer, S., Shu, C., Xi, P., 2011. Landmark-free posture invariant human shape correspondence. Vis. Comput. 27, 843–852.
Yan, P., Bowyer, K.W., 2007. Biometric recognition using 3D ear shape. Pattern Anal. Mach. Intell. 29, 1297–1308.
Yuan, Zh, 1990. Qinshihuang ling bingmayong yanjiu. Cultural Relics Press, Beijing.