Journal of Archaeological Science: Reports 46 (2022) 103676

Contents lists available at ScienceDirect

Journal of Archaeological Science: Reports

journal homepage: www.elsevier.com/locate/jasrep

Can we combine 3D data obtained with a MicroScribe digitising arm and photogrammetry to address bioarchaeological research questions?

Mark Collard [a,*], Keith Dobney [b,c], Kimberly A. Plomp [a,d]

[a] Department of Archaeology, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada
[b] Department of Archaeology, Classics and Egyptology, University of Liverpool, 14 Abercromby Square, Liverpool L69 7WZ, United Kingdom
[c] Department of Archaeology, University of Sydney, The University of Sydney, NSW 2006, Australia
[d] Archaeological Studies Program, University of the Philippines Diliman, Albert Hall Building, Lakandula Street, Diliman, Quezon City 1101, Philippines

ARTICLE INFO

Keywords:
Virtual anthropology
3D shape analysis
Photogrammetry
MicroScribe digitising arm
Facial shape
Crania
Geometric morphometrics

ABSTRACT

Virtual methods for studying human remains are becoming increasingly popular in bioarchaeology, and the rate of technological innovation in the last few years has been such that we now have multiple options to choose from when collecting data. This raises the question of whether datasets generated with different methods are transposable. In the study reported here, we investigated whether it is valid to combine 3D data obtained with a MicroScribe digitising arm and 3D data collected via photogrammetry. We did so by simulating a population-based analysis similar to those commonly undertaken in bioarchaeology. Our sample comprised 19 crania from two ethnic groups, Ancient Egyptians and Guanches, and the landmarks we employed pertained to facial shape.

The analyses yielded several findings. First, we found that photogrammetry was significantly more precise than the MicroScribe digitising arm. Second, the photogrammetry-based method revealed the existence of facial shape differences between the two ethnic groups that were not captured by the MicroScribe-based method. Third, we found that the two methods did not consistently capture the same facial shapes—they did for one of the ethnic groups but not for the other. Fourth, the analyses indicated that using the two methods can result in ethnic group-level differences in facial shape when they are applied to individuals from a single ethnic group. Lastly, the two methods of data collection yielded different patterns of variation in facial shape. Together, these findings suggest that combining 3D landmark coordinates collected with a MicroScribe and those obtained via photogrammetry may introduce considerable error into an analysis, and, consequently, bioarchaeologists should be cautious about doing so.

1. Introduction

Over the last 30 years, bioarchaeologists have increasingly turned to digital methods for studying human remains, so much so that it has been recently argued that these methods are "shaping the future of the discipline" (Ulguim 2018: 191). The rate of technological innovation in the last few years has been such that bioarchaeologists now have multiple options to choose from when they collect and analyse virtual data from human skeletal remains. An important but under-researched question this raises is whether different data-capture methods result in datasets that are transposable. If they are, then it will be possible to compare the results of studies involving different data collection methods directly and to combine datasets with confidence. If not, then comparing studies that deploy different methods of data collection will be more complicated and combining datasets will be more difficult, if not impossible. In the present study, we focus on this issue as it pertains to three-dimensional (3D) shape datasets, i.e., datasets comprising the Cartesian coordinates of landmarks designed to capture the shape of objects such as teeth and bones.

Currently, there are two main approaches to the creation of 3D shape datasets for bioarchaeological research. These approaches are distinguished by whether or not the focal object has to be touched in order to record the X,Y,Z coordinates of a landmark. The contact approach employs a digitising arm, which is a portable electromechanical device that enables the coordinates of a set of landmarks to be recorded on a bone or tooth by touching the object with a stylus. In contrast, the non-contact

* Corresponding author at: Department of Archaeology, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada.
E-mail address: mcollard@sfu.ca (M. Collard).

https://doi.org/10.1016/j.jasrep.2022.103676
Received 8 July 2021; Received in revised form 8 September 2022; Accepted 4 October 2022

approach employs images to create a 3D model of a bone or tooth, and a computer program is then used to record the coordinates of landmarks on the virtual object. There are three main ways of creating the images used in the non-contact approach—laser scanning, computed tomography (CT), and photogrammetry. Among bioarchaeologists, photogrammetry is rapidly becoming the most popular of these non-contact methods. This is because it is relatively cheap, easy to use, and yields high-resolution 3D models. In contrast, laser scanners are relatively slow and often suffer from poor resolution, while CT scanners are not straightforward to access. Few museums have them, and those that are operated by other organisations (e.g. hospitals) are invariably in high demand. It also tends to be expensive to produce CT scans.

Several studies have compared datasets generated with non-contact methods (Butnariu et al., 2013; Katz and Friess 2014; Evin et al. 2016; Jurda and Urbanová 2016; Fruciano et al. 2017). Taken together, the findings of these studies indicate that any differences introduced by the techniques are considerably smaller than the morphological differences between the objects under study (Katz and Friess 2014; Evin et al. 2016). This means that the results of studies in which 3D models have been created with one non-contact method (e.g. laser scanning) can be confidently compared with the results of studies in which another non-contact method has been used to generate 3D models (e.g. photogrammetry). It also means that landmark data recorded on 3D models created with two or more of the non-contact methods can be combined into a single dataset with confidence.

There have also been a number of attempts to compare data collected with a digitising arm to data generated with one or more of the non-contact methods (Pedro 2013; Algee-Hewitt and Wheat 2016; Vu et al. 2017; Adcox 2019; Seguchi et al. 2019; Bertsatos et al. 2020; Waltenberger et al (2021)). The majority of these studies have involved comparing digitising arm-generated data to those obtained from 3D models created with a laser scanner (Pedro 2013; Algee-Hewitt and Wheat 2016; Vu et al. 2017; Seguchi et al. 2019). Generally, these studies suggest that the two types of data are comparable and therefore can be combined without introducing a significant amount of error. However, Seguchi et al. (2019) found that non-homologous landmarks were more difficult to replicate with a digitising arm, which implies that the laser scanner method is a better option when using such landmarks.

To date, only three published studies have compared data generated with a digitising arm to data obtained via photogrammetry. In the earliest of the three studies, Adcox (2019) recorded landmark coordinates on three adult human crania using photogrammetry, a digitising arm, and a number of other methods. He repeated landmark acquisition nine times with each method and compared the variance across the repeated datasets. The analyses suggested that the four methods have a similar level of precision (i.e. a similar level of measurement reproducibility) and concluded from this that data obtained with them can be combined without introducing error. Bertsatos et al. (2020) recorded the x,y,z coordinates of 56 landmarks on 50 adult human crania using photogrammetry and a digitising arm, and then assessed the interchangeability of the data. They found that while the methods had similar levels of precision when a single observer recorded the data, the digitising arm was more prone to interobserver error. Bertsatos et al. (2020) also investigated whether the precision of the methods differed among the three different types of landmarks, Type I landmarks (e.g. bregma, glabella), Type II landmarks (e.g. the tip of a protuberance), and Type III landmarks (e.g. equally spaced landmarks on a curve). Significantly for present purposes, Bertsatos et al. (2020) found that Type I landmarks were more reliably recorded via photogrammetry than with a digitising arm. Most recently, Waltenberger et al. (2021) reported a study that compared the reliability of osteological data collected from four articulated pelves with a CT scanner, a 3D structured light scanner, a photogrammetric system, and a digitising arm. They concluded that data acquisition with a digitising arm is more prone to error than CT scanning, 3D structured light scanning, and photogrammetry. Thus, while Adcox's (2019) findings indicate that landmark data collected with a digitising arm and photogrammetry are comparable, Bertsatos et al.'s (2020) and Waltenberger et al.'s (2021) results suggest they may not be.

The goal of the study reported here was to shed further light on the appropriateness of combining 3D landmark data collected with a digitising arm and 3D landmark data obtained with photogrammetry. The study differed from those of Adcox (2019) and Bertsatos et al. (2020) in that it was designed to be informative regarding the type of population-based analyses of 3D skeletal shape variation that are common in bioarchaeology. Accordingly, a single observer used both methods to capture 3D landmark data from human crania from two different populations. We then sought to answer three questions. The first was, do the data generated with the two methods show a similar level of precision? Second, does combining landmark coordinates obtained with the two methods introduce error into the dataset and, if so, what is the scale of this error? Third, do the two methods of data collection capture the same pattern of morphological differences between different (in this case two) ethnic groups?

2. Materials and methods

The sample comprised 19 crania from two different ethnic groups: Pre-Spanish Guanches from the Canary Islands (n = 11) and Ancient Egyptians (n = 8). Collected in the 19th century, the crania are currently curated at the University of Edinburgh's Anatomical Museum. Individuals were recorded as female or male based on the original collection records (i.e. we did not sex them ourselves). Eight of the individuals were female and 12 were male. Only adult crania were included in the sample in order to avoid the confounding effects of ontogeny; individuals were judged to be adult on the basis of dental eruption.

The specific photogrammetry method we used was outlined by Evin et al. (2016). We took 150 photographs of each cranium with an eight megapixel digital single-lens reflex Canon EOS 77D camera and a 50 mm lens. With the cranium placed on a rotating table, the photographs were shot at intervals of approximately 10°. Following Evin et al. (2016), a 3D scale was employed as a target marker. We then used Agisoft Metashape (Agisoft, 2019) to create a 3D model of the cranium from the

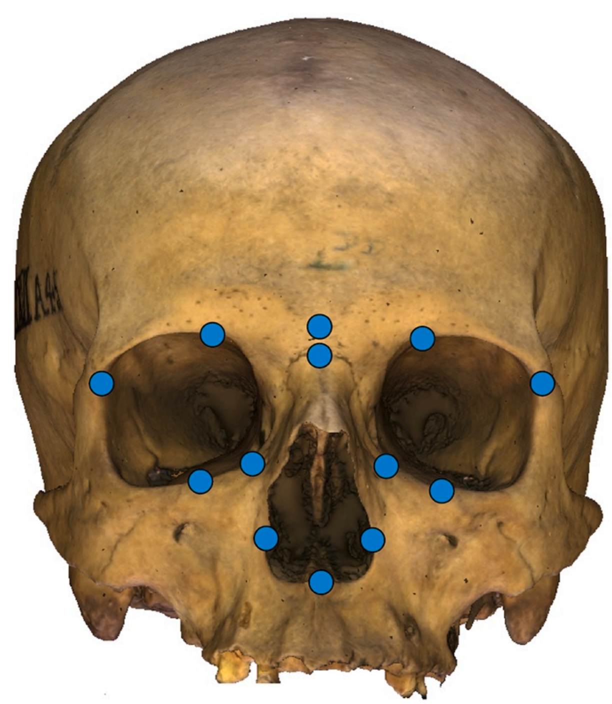

Fig. 1. Location of the 13 landmarks used to capture facial shape.

photographs. We aligned all the photographs with the accuracy level set to 'high', and then produced a depth map, again with the accuracy level set to 'high'. Next, we used the depth map to create a mesh without implementing any additional smoothing techniques. Thereafter, we exported the mesh as a 3D model (.ply). Lastly, the 3D model was imported into MorphoDig (Lebrun 2018) and the 3D Cartesian coordinates of 13 facial landmarks captured twice (Fig. 1). All landmarks were Type I landmarks, according to Bookstein's (1997) widely used scheme. We selected Type I landmarks because they are the most reliable type of landmark (Bookstein, 1997) and thus should maximise the probability of the two methods yielding data that are statistically indistinguishable.

The digitising arm we used was a MicroScribe MLX (https://gomeasure3d.com/microscribe/) with a standard stylus tip. The landmarks we recorded were the same as the ones we captured on the photogrammetry-derived models (Fig. 1). A single observer (KAP) operated the MicroScribe to avoid the problem of inter-observer error. KAP has considerable experience collecting data with a MicroScribe (e.g. Plomp et al. 2013, 2015, 2019a,b, 2020, 2021a,b). As with the photogrammetry, the coordinates of the landmarks were collected twice.

The data used in the study are available from the Dryad Digital Repository (https://doi.org/10.5061/dryad.tmpg4f524).

Once data collection was completed, the two sets of photogrammetry-derived coordinates and the two sets of MicroScribe digitising arm-derived coordinates were combined into a single dataset.

We employed the dataset in five analyses. In the first, we investigated whether the landmarks recorded with the two methods of data collection had the same level of precision. To accomplish this, the entire dataset was subjected to generalised Procrustes analysis (GPA), which converts the Cartesian coordinates for a given landmark configuration into Procrustes coordinates by removing translational and rotational effects and scaling the landmark configuration to centroid size (Slice 2007). Then, for each cranium, we subtracted the second set of photogrammetry-derived coordinates from the first set of photogrammetry-derived coordinates, changed all negative differences to positive ones, and calculated the average difference across all the landmarks. Next, we carried out the same procedure for the two sets of coordinates recorded with the MicroScribe. Subsequently, we calculated a total average difference for the photogrammetry data and a total average difference for the MicroScribe data. Lastly, we used the independent samples *t*-test (two-tailed) to assess the significance of the difference between the two total average difference values.

In the second analysis, we investigated the scale of the shape differences arising from the use of the two data collection methods. First, we averaged the two sets of coordinates obtained with photogrammetry for each ethnic group. We then did the same for the two sets of coordinates obtained with the MicroScribe digitising arm. This meant that each cranium was represented twice in the reduced dataset, once by the averages of the photogrammetry-derived coordinates and once by the averages of the MicroScribe-derived coordinates. In the next step of the analysis, the reduced dataset was subjected to GPA. Subsequently, we created four Operational Taxonomic Units (OTUs): 1) Guanche_photogrammetry (Guanche_P), 2) Guanche_digitising arm (Guanche_DA), 3) Ancient Egyptian_photogrammetry (Egyptian_P), and 4) Ancient Egyptian_digitising arm (Egyptian_DA). Thereafter, we subjected the data to principal components analysis (PCA), retaining only the principal components (PC) that accounted for 5 % or more of the total shape variance (Zelditch et al. 2004). In the last step of the analysis, we subjected the retained PCs to MANOVAs to assess the significance of the differences among the following pairs of OTUs: Egyptian_P *vs* Egyptians_DA, Guanche_P *vs* Guanche_DA, Egyptian_P *vs* Guanche_P, and Egyptian_DA *vs* Guanche_DA. In these comparisons, we calculated the effect sizes (partial eta-squared values) in addition to the standard statistics.

The third analysis also investigated the scale of the shape differences arising from the use of the two data collection methods. The analysis followed the steps carried out in the previous analysis up to the point of creating the four OTUs. After forming the OTUs, we calculated all possible pairwise Procrustes distances among the individuals. Next, we carried out four independent samples *t*-tests (two-tailed). The first *t*-test compared all the pairwise distances among the individuals in the Egyptian_P and Egyptian_DA OTUs to the pairwise distances among the individuals in the Egyptian_P and Guanche_DA OTUs. The second *t*-test compared all the pairwise distances among the individuals in the Egyptian_P and Egyptian_DA OTUs to the pairwise distances among the individuals in the Egyptian_DA and Guanche_DA OTUs. In the third *t*-test, we compared all the pairwise distances among the individuals in the Guanche_P and Guanche_DA OTUs to the pairwise distances among the individuals in the Egyptian_DA and Guanche_DA OTUs. In the fourth *t*-test, we compared all the pairwise distances among the individuals in the Guanche_P and Guanche_DA OTUs to the pairwise distances among the individuals in the Egyptian_P and Guanche_P OTUs. The goal of these *t*-tests was to establish whether the differences introduced by the use of two methods were smaller than, indistinguishable from, or larger than, the differences between the two ethnic groups when a single method was employed.

In the fourth analysis, we sought to ascertain whether the two methods of data collection had captured the same pattern of facial shape differences between the two ethnic groups. The analysis followed the steps carried out in the previous two analyses up to the point of running the PCA. Once the PCA had been completed, we created a scatter-plot depicting the shape variance captured by the first two PCs.

The goal of the fifth and final analysis was also to ascertain whether the two methods of data collection had captured the same pattern of morphological differences between the two ethnic groups. To do this, we divided the dataset into a photogrammetry dataset and a MicroScribe dataset, and then subjected each to GPA. Next, we calculated the Procrustes distance between each individual and the group mean in the photogrammetry dataset, and rank-ordered the individuals according to their distance from the mean. Thereafter, we did the same with the MicroScribe dataset. Lastly, we compared the rank orders yielded by the two datasets.

The GPAs and PCAs were performed in MorphoJ (Klingenberg, 2011), the Procrustes distances in Morphologika (O'Higgins and Jones, 2006), and the *t*-tests in SPSS.

3. Results

The average difference between the two sets of coordinates recorded with the MicroScribe digitising arm was 0.00186, while the average difference for the two sets of coordinates captured with photogrammetry was 0.000000000752. According to the *t*-test, these differences were significant ($t = 3.657$, $F = 4.709$, $p < 0.001$). Thus, the photogrammetry-based method of collecting 3D landmark data was more precise than the MicroScribe-based one.

The results of the second analysis are summarised in Table 1. Only two of the four pairwise MANOVAs returned significant p-values. These were the Egyptian_P *vs* Guanche_P MANOVA, and the Guanche_P *vs* Guanche_DA one. This indicates that the photogrammetry method captured statistically different coordinates for the Ancient Egyptians and the Guanches, revealing the existence of facial shape differences between the two ethnic groups, but the MicroScribe-based method did not.

The results of the pairwise MANOVAs also indicate that there were

Table 1
Results of Procrustes MANOVAs comparing the OTUs, ordered by effect size. λ = sum of squares. F = F value. η^2p = partial eta-squared.

OTU comparisons	λ	F	p-value	η^2p
Egyptian_P *vs* Egyptian_DA	0.764	0.462	0.820	0.236
Egyptian_DA *vs* Guanche_DA	0.406	1.217	0.424	0.594
Guanche_P *vs* Guanche_DA	0.404	3.690	0.019	0.596
Egyptian_P *vs* Guanche_P	0.308	4.486	0.013	0.692

significant differences between the coordinates produced by the two methods when they were applied to the Guanche sample. According to the partial eta values, these differences were larger than the differences between the coordinates for the two ethnic groups when a single method was used. Thus, not only did the two methods not capture the same facial shapes for one of the samples, but also the differences between the coordinates produced by the two methods for one of the groups were within the range of differences between the two ethnic groups when a single method was employed. The latter finding means that using the two methods can result in ethnic group-level differences in facial shape even when they are applied to individuals from a single ethnic group.

Table 2 summarises the results of the third analysis, which also sought to assess the scale of the shape differences arising from the use of the two data-collection methods. Three of the *t*-tests returned insignificant p-values (Egyptian_P and Egyptian_DA *vs* Egyptian_P and Guanche_P, Egyptian_P and Egyptian_DA *vs* Egyptian_DA and Guanche_DA, and Guanche_P and Guanche_DA *vs* Egyptian_DA and Guanche_DA) while one returned a significant p-value (Guanche_P and Guanche_DA *vs* Egyptian_P and Guanche_P). The key results for present purposes are the insignificant ones. The Egyptian_P and Egyptian_DA *vs* Egyptian_P and Guanche_P *t*-test and the Egyptian_P and Egyptian_DA *vs* Egyptian_DA and Guanche_DA *t*-test indicate that the differences introduced by the use of two methods were statistically indistinguishable from the differences between the two ethnic groups when a single method was employed. This confirms that using the two methods can produce ethnic group-level differences in facial shape when they are applied to individuals from a single ethnic group.

Six PCs were generated in the fourth analysis. The first and second of these accounted for 27 % and 11 % of total shape variance, respectively. When they were plotted against each other, it was clear that the patterns of shape variation produced by the two data collection methods were different (Fig. 2). Most notably, the landmark configurations collected using the MicroScribe digitising arm tended to score more positively on PC2 than those obtained with photogrammetry.

Table 3 summarises the results of the fifth analysis. The second column of the table lists the ranks of the individuals based on the photogrammetry-derived Procrustes distances between the crania and the group mean, while the third column lists the ranks of the individuals based on the MicroScribe-derived Procrustes distances between the crania and the group mean. Only three individuals have the same rank in both columns, SK 45, SK 50, and SK 31. The ranks of the remaining individuals differ depending on which method was used to collect the data. As such, the results of the fifth analysis also indicate that the two data-collection methods did not capture the same pattern of morphological differences between the two ethnic groups.

4. Discussion

In the present study, we investigated whether it is appropriate to combine 3D coordinates obtained with a MicroScribe digitising arm and 3D coordinates produced through photogrammetry for the purposes of bioarchaeological research. We did so by simulating an ethnic group-based comparative analysis similar to those commonly undertaken in bioarchaeology. Our sample comprised crania from two ethnic groups, Ancient Egyptians and Guanches, and the landmarks we employed pertained to facial shape.

Our analyses indicated that there was a significant difference in the level of precision between the two methods, with the photogrammetry-based method being more precise than the MicroScribe-based method. We also found that the photogrammetry-based method revealed the existence of facial shape differences between the two ethnic groups, but the MicroScribe-based method did not. Additionally, our analyses revealed that the two methods of data collection did not consistently capture the same facial shapes—they did for one of the ethnic groups but not for the other. A fourth important finding was that using the two methods can result in ethnic group-level differences in facial shape even when they are applied to individuals from a single ethnic group. Lastly, we discovered that the two methods of data collection yielded different patterns of variation in facial shape.

In the Introduction, we explained that three previous studies have compared digitising arm-derived 3D data photogrammetry-derived 3D data, with a view to determining whether such data can be combined—Adcox (2019), Berstatos et al. (2020), and Waltenberger et al. (2021). Adcox (2019) generated data with several methods, including a digitising arm and photogrammetry. He found that different datasets had a similar level of precision and concluded that it was valid to combine them for the purposes of analysis. Our results are clearly inconsistent with this. We suspect the reason we obtained different results from Adcox's (2019) is that his sample was markedly smaller than our one (three specimens versus 19).

The situation is different with Berstatos et al.'s (2020) and Waltenberger et al.'s (2021) studies. As we noted earlier, Berstatos et al. (2020) found that the digitising arm and photogrammetry methods had similar levels of precision when a single observer recorded the data, but the digitising arm method was more prone to interobserver error. They also found that the precision of the methods differed when different types of landmarks were utilised. Specifically, they found that the photogrammetry method was more precise than the digitising arm method when only Type I landmarks were employed. Given that our landmarks were all Type I landmarks, our results are most comparable to Berstatos et al.'s (2020) when they examined the precision of the different types of landmarks. And that means the two studies are in agreement. They both indicate that the photogrammetry method is more precise than the digitising arm method when recording Type I landmarks.

To reiterate, Waltenberger et al. (2021) compared the reliability of osteological data collected from four articulated pelves with a CT scanner, a 3D structured light scanner, a photogrammetric system, and a digitising arm. They found that data acquisition with a digitising arm is more prone to error than CT scanning, 3D structured light scanning, and photogrammetry. So, our results are also consistent with theirs.

That we, Berstatos et al. (2020, and Waltenberger et al. (2021) found the photogrammetry method to be more precise than the digitising arm method is perhaps unsurprising in retrospect. A number of factors that can affect landmark acquisition with a digitising arm are not a concern when obtaining landmarks with the photogrammetry method. These include accidental movement of the table and/or the crania, and a shaky hand. In our experience, even a door being forcefully closed nearby can result in minor vibrations that can impact placement of the tip. In addition, Microscribes should be calibrated on a regular basis and it is our impression that this is rarely done, probably because it entails sending the device to a US-based company, which is costly and time-consuming.

Regardless of the cause of the difference in precision between the photogrammetry and digitising arm method, its existence means that combining landmark data obtained with the two methods of data collection has a high probability of introducing error that is of sufficient magnitude to affect the findings of a bioarchaeological study. Given this,

Table 2
Results of independent samples *t*-tests (two tailed) comparing average pairwise Procrustes distances among individuals assigned to the four OTUs.

OTU comparisons	*t*	p-value
Egyptian_P and Egyptian_DA *vs* Egyptian_DA and Guanche_DA (average pairwise Procrustes distances: 0.0973 *vs* 0.0106)	−0.970	0.333
Egyptian_P and Egyptian_DA *vs* Egyptian_P and Guanche_P (average pairwise Procrustes distances: 0.0973 *vs* 0.105165)	0.487	0.627
Guanche_P and Guanche_DA *vs* Egyptian_DA and Guanche_DA (average pairwise Procrustes distances: 0.104506 *vs* 0.0106)	−0.319	0.750
Guanche_P and Guanche_DA *vs* Egyptian_P and Guanche_P (average pairwise Procrustes distances: 0.104506 *vs* 0.105165)	−2.665	0.008

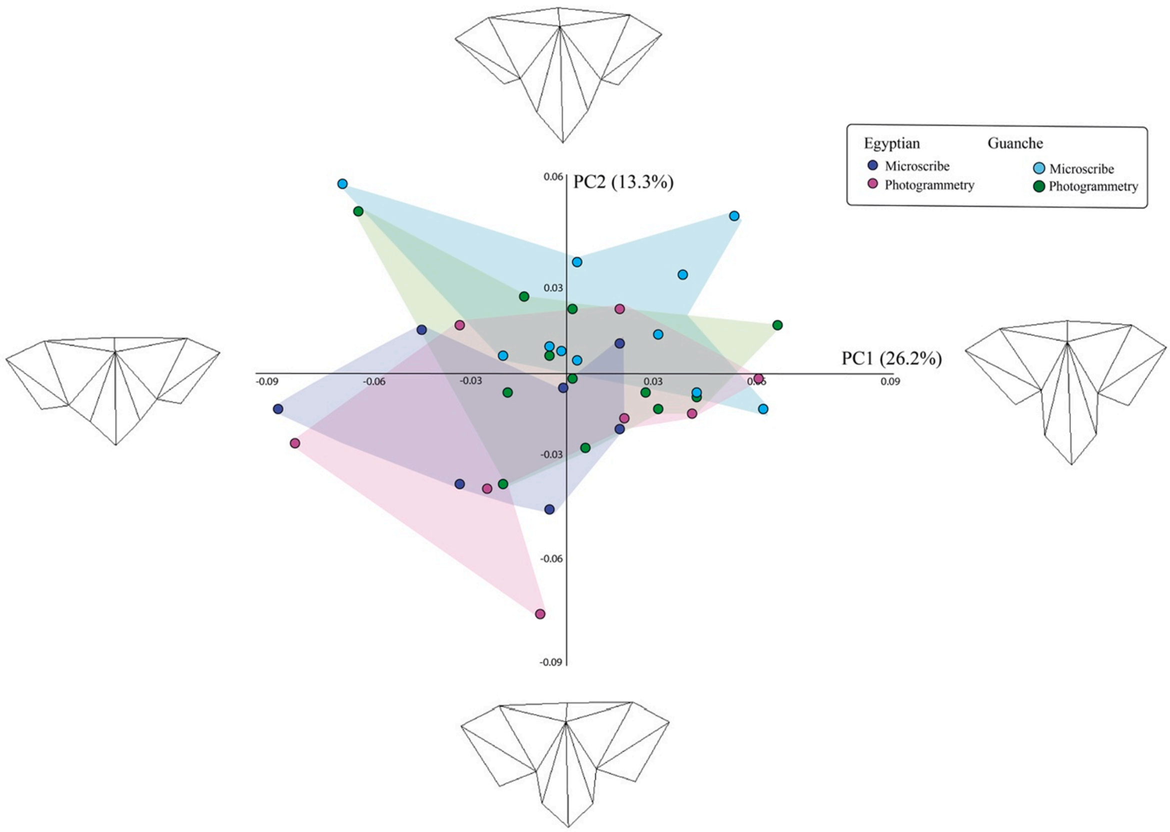

Fig. 2. PCA scatterplot depicting the shape variance of the sample when PC1 and PC2 are plotted against each other.

Table 3
Ranking of the Procrustes distances between each individual cranium and the mean based on method employed.

Rank order	Photogrammetry	MicroScribe digitising arm
1*	Guanche - SK 45	Guanche - SK 45
2	Guanche - SK 46	Guanche - SK 47
3	Guanche - SK 47	Ancient Egyptian - SK 8
4	Ancient Egyptian - SK 10	Ancient Egyptian - SK 9
5	Ancient Egyptian - SK 7	Ancient Egyptian - SK 18
6	Guanche - SK 48	Guanche - SK 57
7	Ancient Egyptian - SK 9	Guanche - SK 48
8	Ancient Egyptian - SK 30	Guanche - SK 46
9	Guanche - SK 61	Ancient Egyptian - SK 30
10	Ancient Egyptian - SK 13	Ancient Egyptian - SK 7
11	Guanche - SK 53	Guanche - SK 63
12	Guanche - SK 57	Ancient Egyptian - SK 10
13	Guanche - SK 63	Guanche - SK 62
14	Ancient Egyptian - SK 18	Ancient Egyptian - SK 13
15*	Guanche - SK 50	Guanche - SK 50
16	Ancient Egyptian - SK 8	Guanche - SK 61
17	Guanche - SK 62	Guanche - SK 52
18	Guanches - SK 52	Guanche - SK 53
19*	Ancient Egyptian - SK 31	Ancient Egyptian - SK 31

The distances are organised in ascending order, with the crania closest to the mean at the top of the list and the crania farthest from the mean at the bottom of the list. * = same rank order produced with photogrammetry and MicroScribe digitising arm.

we suggest that combining datasets of 3D landmark coordinates obtained with a digitising arm and photogrammetry should be avoided, or at least approached very cautiously. It could be appropriate to combine such datasets if one has approximately equal numbers of digitising arm-derived and photogrammetry-derived data points in each OTU (e.g. ethnic group), because the method-related errors can be expected to be random across the OTUs and therefore will not inflate the probability of obtaining a significant result. However, one should avoid comparing OTUs when there is a substantial imbalance in the number of digitising arm-derived and photogrammetry-derived data points in the OTUs. In such a situation, there is a strong possibility that the method-related errors will give rise to significant but false results.

The finding that the photogrammetry method is more precise than the digitising arm method has implications beyond the question of whether data obtained with the two methods can be safely combined for bioarchaeological research. Most obviously, it argues in favour of the use of photogrammetry in preference to a digitising arm, if one has a choice. It should be noted that the precision of photogrammetry can be expected to vary depending on the size and complexity of the objects being analysed. For example, the precision of data obtained from a photogrammetry-derived model of a human cranium will likely be higher than that of data obtained from a much smaller, more complex object, such as an individual sphenoid. However, there are some ways to increase the precision of photogrammetry, such as using more than one 3D marker target to help with photograph alignment, and using a larger number of photographs to create the model. Needless to say, this may not always be possible due to time constraints (e.g. using more photographs will increase processing time), issues with access to equipment, etc. In such cases, we suggest repeating landmark acquisition two to three times for each specimen and then using the average coordinates in subsequent analyses. This should reduce the impact of errors introduced by the method of data collection.

With regard to future research, it would be helpful to repeat our study with crania that belong to different ethnic groups. We found that the differences between the two methods were greater for one ethnic group than for the other, and this raises the possibility that there may be some groups where combining data generated with the two methods is appropriate. We doubt this is likely to be the case, but it is worth further investigation. It would also be useful to combine our research protocol with landmarks for another region of the cranium and/or a different region of the skeleton such as the 1st lumbar vertebra or femur, since it is feasible that the difference in performance between the two methods varies by region and we, by chance, selected a region for which the digitising arm performs particularly poorly. Lastly, it would be helpful to repeat our study with Type II and Type III landmarks as well as Type I landmarks. Given that Type I landmarks are generally considered to be more reliable than Type II and Type III landmarks (Bookstein, 1997), the fact that Berstatos et al. (2020) found a difference in precision between the two methods of data collection when Type I landmarks were used but not when Type II and Type III landmarks were employed is surprising. Repeating our study with landmarks of all three types would shed light on whether Berstatos et al.'s (2020) finding regarding Type II and Type III landmarks was reliable or fortuitous.

5. Conclusions

The study reported here addressed an important but under-researched question—namely, are 3D data collected with different methods transposable for the purposes of bioarchaeological research? Together, the results of the study suggest that the answer to this question is 'no'. Combining datasets of 3D landmark coordinates obtained via photogrammetry and 3D landmark coordinates collected with a MicroScribe should be done cautiously. Most significantly, one should avoid employing combined datasets in studies that involve comparing OTUs if there is a substantial imbalance in the number of photogrammetry-derived and MicroScribe-derived data points between the OTUs (i.e. one OTU mostly has photogrammetry-derived data and another has mostly MicroScribe-derived data).

CRediT authorship contribution statement

Mark Collard: Conceptualization, Supervision, Formal analysis, Resources, Writing – original draft, Writing – review & editing. **Keith Dobney:** Supervision, Writing – review & editing. **Kimberly A. Plomp:** Conceptualization, Formal analysis, Investigation, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data have been made available on Dryad (doi:10.5061/dryad.tmpg4f524)

Acknowledgements

We thank Malcolm MacCallum of the University of Edinburgh's Anatomical Museum for providing access to the crania. We also thank the editor and two anonymous reviewers for their feedback, which markedly improved the paper. We are grateful for their time, trouble, and expertise. Our research was supported by the European Research Council's Marie Skłodowska-Curie Actions program (Horizon 2020 - 748200), the Social Sciences and Humanities Research Council of Canada (895-2011-1009), the Canada Research Chairs Program (228117 and 231256), the Canada Foundation for Innovation (203808), the British Columbia Knowledge Development Fund (862-804231), MITACS (IT03519), the Wenner-Gren Foundation (62447), Simon Fraser University (14518), and the University of Liverpool.

References

Adcox D. 2019. The utility of digital imaging technologies for the virtual curation and metric analysis of skeletal remains. Middle Tennessee State University: Honors College Thesis. https://jewlscholar.mtsu.edu/handle/mtsu/6089.
Algee-Hewitt, B., Wheat, A., 2016. The reality of virtual anthropology: comparing digitizer and laser scan data collection methods for the quantitative assessment of the cranium. Am. J. Phys. Anthropol. 160, 148–155.
Bertsatos, A., Gkaniatsou, E., Papageorgopoulou, C., Chovalopoulou, M.E., 2020. "What and how should we share?" An inter-method inter-observer comparison of measurement error with landmark based craniometric datasets. Anthropol. Anz. 77, 109–120.
Bookstein, F., 1997. Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. Med. Image Anal. 1, 225–243.
Butnariu, S., Gîrbacia, F., Orman, A., 2013. Methodology for 3D reconstruction of objects for teaching virtual restoration. Int. J. Comput. Sci. 3 (01), 16–21.
Evin, A., Souter, T., Hulme-Beaman, A., Ameen, C., Allen, R., Viacava, P., Larson, G., Cucchi, T., Dobney, K., 2016. The use of close-range photogrammetry in zooarchaeology: creating accurate 3D models of wolf crania to study dog domestication. J. Archaeolog. Sci.: Rep. 9, 87–93.
Fruciano, C., Celik, M., Butler, K., Dooley, T., Weisbecker, V., Phillips, M., 2017. Sharing is caring? Measurement error and issues arising from combining 3D morphometric datasets. Ecol. Evol. 7, 7034–7046.
Jurda, M., Urbanová, P., 2016. Three-dimensional documentation of Dolni Vestonice skeletal remains: can photogrammetry substitute laser scanning? Anthropologie 54, 109–118.
Katz, D., Friess, M., 2014. Technical Note: 3D from standard digital photography of human crania – a preliminary assessment. Am. J. Phys. Anthropol. 154, 152–158.
Klingenberg, C.P., 2011. MorphoJ: an integrated software package for geometric morphometrics. Mol. Ecol. Resour. 11, 353–357.
Lebrun R. 2018. MorphoDig, an open-source 3D freeware dedicated to biology. IPC5, Paris, France.
O'Higgins, P., Jones, N., 2006. Tools for statistical shape analysis. Hull York Medical School. http://sites.google.com/site/hymsfme/resources.
Pedro, A.S.P., 2013. Brain landmarks and paleoneurology: comparing physical and laser scan endocasts in living hominoids. University of Coimbra, Dissertação de Mestrado em Evolução e Biologia Humanas.
Plomp, K.A., 2013. Quantifying Palaeopathology Using Geometric Morphometrics. Durham University. PhD Dissertation.
Plomp, K.A., Roberts, C.A., Strand, V.U., 2015. Morphological characteristics of healthy and osteoarthritic joint surfaces in archaeological skeletons. Int. J. Osteoarchaeol. 25, 515–527.
Plomp, K.A., Strand Viðarsdóttir, U., Weston, D., Dobney, K., Collard, M., 2019a. Potential adaptations for bipedalism in the thoracic and lumbar vertebrae of *H. sapiens*: A 3D comparative analysis. J. Hum. Evol. 137, 102693.
Plomp, K.A., Strand Viðarsdóttir, U., Weston, D., Dobney, K., Collard, M., 2019b. 3D shape analyses of extant primate and fossil hominin vertebrae support the Ancestral Shape Hypothesis for intervertebral disc herniation. BMC Evol. Biol. 19, 226.
Plomp, K.A., Dobney, K., Collard, M., 2020. Spondylolysis and spinal adaptations for bipedalism: The Overshoot Hypothesis. Evol. Med. Public Health 2020 (1), 35–44.
Plomp, K.A., Dobney, K., Collard, M., 2021a. A 3D basicranial shape-based assessment of local and continental northwest European ancestry among 5th to 9th century CE Anglo-Saxons. PLoS ONE 16 (6), e0252477.
Plomp, K.A., Dobney, K., Gestdóttir, H., Price, N., Collard, M., 2021b. The composition of the founding population of Iceland: a basicranial perspective. PLoS ONE 16 (2), e0246059.
AgiSoft Metashape Professional (Version 1.7.2). 2019. Retrieved from http://www.agisoft.com/downloads/installer/.
Seguchi, N., Dudzik, B., Murphy, M.M., Prentiss, A.M., 2019. Introduction. In: Seguchi, N., Dubzik, B. (Eds.), 3D Data Acquisition for BioArchaeology, Forensic Anthropology, and Archaeology. Elsevier Academic Press, London.
Slice, D., 2007. Geometric morphometrics. Ann. Rev. Anthropol. 36, 261–281.
Ulhuim, P., 2018. Models and metadata: The ethics of sharing bioarchaeoloical 3D models online. Archaeologies 14, 189–228.
Vu, A., Chundury, R., Perry, J., 2017. Comparison of the FaroArm laser scanner with the Microscribe digitizer using basicranial measurements. J. Craniof. Surg. 28, e460–e463.
Waltenberger, L., Rebay-Salisbury, K., Mitteroecker, P., 2021. Three-dimensional surface scanning methods in osteology: A topographical and geometric morphometric comparison. Am. J. Phys. Anthropol. 174, 846–858.
Zelditch, M.L., Swiderski, D.L., Sheets, D.H., Fink, W.L., 2004. Geometric morphometrics for biologists: a primer. Elsevier Academic Press, London.