Forensic Science International 242 (2014) 177–185

Contents lists available at ScienceDirect

Forensic Science International

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

Technical note

Accuracy of standard craniometric measurements using multiple data formats

Adam H. Richard [a], Connie L. Parks [a], Keith L. Monson [b,*]

[a] Counterterrorism Forensic Science Research Unit, Visiting Scientist Program, Federal Bureau of Investigation Laboratory Division, Quantico, VA, USA
[b] Counterterrorism Forensic Science Research Unit, Federal Bureau of Investigation Laboratory Division, 2501 Investigation Parkway, VA 22135, USA

ARTICLE INFO

Article history:
Received 8 April 2014
Received in revised form 11 June 2014
Accepted 17 June 2014
Available online 25 June 2014

Keywords:
Craniometrics
Measurement validation
Measurement error
Medical imaging
Laser scanning

ABSTRACT

With continuing advancements in biomedical imaging technologies, anthropologists are increasingly making use of data derived from indirect measurement and analysis of skeletal material. To that end, the purpose of this study was to test the reliability of 26 standard craniometric measurements routinely utilized in forensic casework across several different imaging technologies. Measurements from five crania of known individuals were collected in duplicate by two anthropologists via computed tomography (CT) scans and three-dimensional (3D) laser scans of the known skulls. The laser scans were also used to create prototype models of the known skulls. These prototypes were, themselves, laser-scanned, and measurements were also collected from the prototypes and the laser scans of the prototypes. Measurement sets from each technology were then compared with one another using the previously collected osteometric measurements taken on the crania themselves as the ground truth.

Results indicate that, while the majority of measurements showed no significant differences across data formats, a handful were found to be problematic for particular technologies. For instance, measurements taken in a supero-inferior direction (e.g., BBH, OBH) from CT scans were prone to greater deviation from direct measurements of the cranium than other technologies, especially for CT scans taken at 5 mm thickness and increment. Also, several measurements defined by Type 1 landmarks, particularly those occurring at complicated or indistinct suture junctures (e.g., ASB, ZMB), were found to have high variance across all technologies while measurements based on Type 3 landmarks proved to be highly reproducible. This is contrary to measurements taken directly on crania, in which measures defined by Type 1 landmarks are typically the most reliable, likely attributable to diminished or totally obscured suture definition in the scan data. If medical imaging data are to be increasingly utilized in anthropological studies, it may be prudent to bear in mind that the reliability of measurements taken on an actual skull may not be the same as for measurements taken from medical scans.

1. Introduction

As biomedical scanning technologies have become less expensive and more readily available, a growing body of anatomical data has become available electronically for use in research and medicolegal pursuits. To properly utilize this burgeoning new resource, care must be taken to ensure that measurements collected from these new forms of data are comparable with the physical measurements collected from anatomical specimens that have been used to develop current forensic methods and procedures. A number of studies have been conducted testing both the accuracy and precision of craniofacial measurements using a variety of imaging technologies, including computed tomography (CT) scanners, cone beam computed tomography (CBCT) scanners, digitizers, and laser scanners [1–14]. Some of these studies have sought to evaluate the intrinsic accuracy of the imaging technologies; these studies typically involved placement of some type of marking device prior to scanning in order to eliminate error associated with variability in repeated attempts to locate landmarks for measurement [1–6]. Other studies have focused on the ability of practitioners to produce accurate and precise measurements using these new technologies [7–10]. Still others have looked at the influence that various scanning protocols and procedures can have on measurement accuracy and precision [11–14].

While a number of imaging technologies have been independently tested, few have examined measurement error in the absence of visible landmark indicators across multiple data

* Corresponding author. Tel.: +1 703 632 7847.
E-mail address: keith.monson@ic.fbi.gov (K.L. Monson).

http://dx.doi.org/10.1016/j.forsciint.2014.06.015
0379-0738/Published by Elsevier Ireland Ltd.

formats using the same reference sample. Such information could provide useful guidance in the design of future research studies using nontraditional cranial data. For instance, it may be possible to identify specific measurements that are less reliable when using data from a particular imaging technology or one imaging technology that can be relied upon to provide more accurate and/or precise measurements in general. To explore these possibilities, a research study was designed to compare measurements collected directly from a sample of skulls with measurements taken from CT scans, laser scans, prototype models, and laser scans of prototype models of the same skulls.

2. Methods

Crania of five individuals from the William M. Bass Donated Skeletal Collection were selected for inclusion in the study based on the prior acquisition of CT and laser scans of these crania, as well as the creation of prototypes of the crania based on these laser scans, for previous research in facial approximation [15]. Craniometrics collected from the original skulls were provided courtesy of the University of Tennessee, Knoxville. These data consisted of standard cranial measurements [16] collected for all donations to the Bass collection but do not include mandibular measurements. Three of the five crania had been CT-scanned using 2.5 mm increment and scan slice thickness, while the remaining two were scanned at 5 mm increment and slice thickness. All CT scanning was conducted using a GE Light Speed CT System with 16 slices, helical scan type, and 0.8 s rotation time (General Electric Healthcare, Waukesha, WI). All five crania were also scanned with a FARO Scanarm laser scanner, which is accurate to ±0.356 mm (FARO, Lake Mary, FL). Scan data from the FARO Scanarm was then processed into 3D image files using Polyworks software (InnovMetric Software Inc. Quebec, QC Canada), and these image files were used to create prototype models of the skulls using either an SLA-5000 Stereo Lithography machine or a Projet HD 3000 3D Printer (3D Systems, Inc., Rock Hill, SC). These prototype models were subsequently rescanned using a NextEngine Desktop 3D Scanner, which is accurate to ±1.27 mm (NextEngine, Santa Monica, CA). Measurements were collected from CT and laser scans using tools in the Analysis Module of Materialise's Mimics v. 14 software (Materialise, Plymouth, MI). Digital spreading and sliding calipers were used for the collection of measurements from prototype models (Paleo-Tech Concepts, Crystal Lake, IL).

Two anthropologists each collected a total of 26 craniometric measurements (Table 1) in duplicate for each technology. No mandibular measurements were collected owing to a lack of osteometric data for these measurements for the five subjects selected for the study. Measurements from a given subject were collected at least two weeks, and in many cases several months, apart to prevent replication of exact landmark or caliper placement. Caliper measurements were automatically entered into an Excel spreadsheet via the use of a foot pedal, and measurements collected using Mimics were automatically calculated based on landmark placement and saved for later export into the spreadsheet after all measurements had been collected. These precautions minimized the possibility of measurement bias from manually recording (and therefore knowing) the measurement values as they were being collected. Once all the measurements had been collected, their values were compared with the corresponding direct measurements (treated as the "ground truth" measurements) recorded at the University of Tennessee. Values for intra- and interobserver error were also calculated.

3. Results

Tables 2 and 3 show the summary statistics for intra- and interobserver error, respectively, for each technology. Intraobserver error was calculated by subtracting Trial 2 from Trial 1 for each measurement from both observers. Mean intraobserver differences proved to be less than 0.5 mm for all technologies with standard deviations less than 2.5 mm. Interobserver error was determined by subtracting Observer 2's results from Observer 1's results for each measurement and each Trial. Interobserver differences were even smaller with means less than 0.1 mm and standard deviations falling under 2.5 mm.

To examine the potential differences among the tested technologies, differences from the osteometric data were calculated for each individual measurement that was collected during the study. Osteometric values were subtracted from the collected measurements, so that negative differences would indicate a smaller value relative to the osteometric measurement. Table 4 contains the means and standard deviations of these values for each technology. Measurements from the prototype models had the lowest mean difference from the osteometric values at <0.01 mm and the smallest standard deviation at 1.52 mm. CT scans demonstrated the highest mean difference from osteometric values at 0.25 mm, while measurements from laser scans of the prototype models had the greatest standard deviation at 2.79 m.

Box-and-whisker plots (Figs. 1–4) help to visualize the variation of the individual measurement errors and to identify measurements that may be less consistent for a certain technology. Measurements PAC, OCC, ASB, and ZMB have higher measurement variance across all tested technologies, although ASB was quite good for the prototype models. Additionally, FRC had higher variance for the prototype and laser of prototype trials. Another measurement that stands out in these plots is UFHT. Only two of the skulls used in this study included osteometric measurements for UFHT, and both of these skulls exhibited severe alveolar

Table 1
Collected measurements.

Measurement	Abbreviation	Landmarks	Measurement	Abbreviation	Landmarks
Max. cranial length	GOL	g–op	Nasal breadth	NLB	al–al
Max. cranial breadth	XCB	eu–eu	Orbital breadth	OBB	d–ec
Bizygomatic breadth	ZYB	zy–zy	Orbital height	OBH	NA
Basion–Bregma height	BBH	ba–b	Biorbital breadth	EKB	ec–ec
Cranial base length	BNL	ba–n	Interorbital breadth	DKB	d–d
Basion–Prosthion length	BPL	ba–pr	Frontal chord	FRC	n–b
Maxillo–Alveolar breadth	MAB	ecm–ecm	Parietal chord	PAC	b–l
Maxillo–Alveolar length	MAL	pr–alv	Occipital chord	OCC	l–o
Biauricular breadth	AUB	au–au	Foramen magnum length	FOL	ba–o
Upper facial height	UFHT	n–pr	Foramen magnum breadth	FOB	NA
Min. frontal breadth	WFB	ft–ft	Mastoid length	MDH	NA
Upper facial breadth	UFBR	fmt–fmt	Biasterionic breadth	ASB	as–as
Nasal height	NLH	n–ns	Zygomaxillary breadth	ZMB	zm–zm