Identification from images: Theory and methods.

A thesis submitted by

Teghan Lucas

Bachelor of Health Sciences (Honours)

Anatomical Sciences

For the degree

Doctor of Philosophy

School of Medicine

Anatomy and Pathology

The University of Adelaide

January 2016

Table of Contents

Abstract

The use of images for the identification of criminals is becoming more prevalent with the increased use of video surveillance systems. Any anatomical trait that is visible on an image could be used to identify an individual, as long as its usefulness as a biometric indicator is known and can be accurately measured.

The mug shot, which was introduced in 1879 by Alphonse Bertillon was the first photograph used in forensic identification from images and since then the human face has been the focus for identification and recognition. However, the usefulness of the face or any other part of the body that could be measured from an image has not been thoroughly investigated.

Population frequencies of various traits are known. However, many studies which investigate the frequencies of traits, use categorical scales of measurement. Categorical scales of measurement have been used to describe the human face and body for centuries, it is not a new technique. The advantages of using categorical scales to describe various anatomical features is, that it is inexpensive to study and does not require specialised technology. As long as an individual is well trained with sufficient knowledge of the human body, categorical scales are generally accepted as a means of describing human variation. The use of categories for description of the human body is currently accepted for research purposes and cases of skeletal identification. However, the use of categories is questioned when describing an individual from an image.

A possible reason for this could be that in image analyses the traits are often too small to see, they are covered by clothing (such as those of the face by a balaclava) or they are subject to image distortion. Therefore, statements made by an expert witness in court proceedings regarding descriptions of anatomical features using categorical scales from images can often be questioned as it is primarily opinion based evidence.

Morphological analyses which use categories for image analyses have been labelled as ‘unreliable’ for the reasons stated above. Much research has concentrated on using interval scales of measurement on anatomical features seen in images. Using metric measurements of images is an attempt to make image identification more reliable by removing the ‘opinions’ of expert witnesses. The methods which are used currently to take measurements from images are time consuming, tedious, have unacceptable error rates and are often expensive.

Increased use of images for identification that will be used as evidence in court cases lead to the establishment of standards by which scientific evidence can be accepted by courts. These standards require the evidence provided by expert witnesses to be reliable, repeatable, peer reviewed and to have known error rates. The only way to make image-based evidence reliable and repeatable is to use interval scales of measurement and to minimize errors.

This thesis proposes that humans are singular in their overall surface anatomy. Therefore the use of interval scales to measure anatomical features for identification from images is justified as a biometric tool. Various methods have been proposed to take reliable measurements from images and to identify the associated error rates.

In order to accomplish this, several investigations were carried out, where each was concerned with a different issue that was involved with the reliable identification of individuals from images.

The first analysis considered whether or not measurements of the human body can be taken from images with precision, regardless of wearing clothes. Light clothing did not affect accuracy of measurements. Bulky and patterned clothing produced greater inaccuracies, but the overall accuracy rate remained at 96%. It was also found that anatomists had the ability to locate anthropometric points with greater precision than the specialists in image analysis.

The second analysis considered the development of a method which could be used in forensic identification to establish the similarities or differences between individuals when large numbers of samples are available (n=3982). The method involves searching for duplicate individuals within a large database and once individuals did not match with another on anthropometric measurements, then they are considered ‘singular’. The term singularity was introduced, as it cannot be debated in a court of law, being a method that could be tested compared to ‘uniqueness’ that is universal. Measurements of the human face were examined to evaluate the value of the method in the identification of an individual. Results showed that the probability of finding two individuals with the exact same eight facial measurements is 1 in a trillion. Thus this is comparable with fingerprints.

The third analysis used the method proposed in the second analysis to investigate the value of body measurements as well as measurements of the face in the identification of an individual. Measurements of the body were compared with those of the face to examine, which measurements were better for the identification of an individual. Results showed that measurements of the body are superior to those of the face with a probability of 1 in a quintillion of finding two “duplicate individuals”. This exceeds the probabilities associated with measurements of the face and is comparable with fingerprint and DNA analyses.

The fourth analysis investigated the effect that measurement errors have in analyses of large anthropometric datasets. In order to achieve this, a formula was developed which converted standard metric units to ‘units of TEM’ (technical error of measurement) and incorporated the measurement errors into reported values. Two large datasets were used, ANSUR (n=3982) and The National Size and Shape Survey of Australia (n =1265). Three examples were used to illustrate the application of the formula: i.e. in forensic investigations, garment construction and study of biological variation. In all examples, using units of TEM was superior to using standard metric units, as it removed inevitable adverse effects that measurement errors have on data.

The final investigation showed that body proportions were not a reliable method for the identification of individuals from images. The error rates associated with the body proportional measurements were equal to the biological variation of individuals.

The information gathered from these five experiments indicates that surface anatomy is sufficient as a biometric tool, which could be applied to identification of individuals from images. Findings in these investigations show that measurements can successfully be taken from images. However, more work needs to be done within the field to reduce error rates.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent for this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Teghan Lucas

Date

Acknowledgements

I would firstly like to thank my supervisors, Professor Maciej Henneberg and Dr. Jaliya Kumaratilake for all their time and effort spent helping me with my research and writing. It is not only the help you provided during my PhD candidature but all the steps leading up to it, I can honestly say that, without you, none of this would have been possible.

I would like to acknowledge Dr Tony Scoleri for assisting with the computer analysis side of things and for your significant input in the first publication, your help is much appreciated.

I would like to acknowledge Emeritus Professor John Fryer, who spent a great deal of time communicating with me, helping me and photocopying huge amounts of resources that I would not have access to otherwise.

I would like to thank both Tavik Morgenstern and Michael Brockhouse for continual help with photographs, image construction for publication and tech support.

I acknowledge Mark Kennedy, Joel Sabine and the security team at The University of Adelaide who gave me access to designated security cameras to conduct my research.

I would like to acknowledge all of my participants who gave up their time to help me with data collection for three of the manuscripts included in this thesis.

I would like to thank all members of the Australasian Society for Human Biology (ASHB) who in 2014 gave myself and others of the Biological Anthropology and Comparative Anatomy Research Unit the opportunity to organise and host the 28th annual ASHB conference. I learnt valuable skills while organising this conference and it is an experience I continue to learn from.

My research has been supported by the Faculty of Health Sciences Divisional Scholarship, School of Medicine postgraduate Travel award and the University of Adelaide.

I would like to thank members of the Biological and Comparative Anatomy Research Unit for their support and scientific conversations throughout my entire PhD. This includes: Malcolm Brinn, Arthur Saniotis, Stella Ioannou, Caitlin Humphrey, Aaron Hermann, Todd England, Pen You Wenpeng, Dante Roccisiano and Arjun Burlakoti.

With special thanks to Amrita Dhugga and Kara Holloway, who provided the perfect balance between distraction and work, making me a more productive person.

Lastly, I would like to thank all the people in my life who put up with me and supported me throughout my academic career thus far: my family, friends, loving partner Uriah and son, Pugsley. You all kept me sane, focused and motivated.

I thank you all.

Manuscripts included in this thesis:

1. Scoleri, T, Lucas, T, Henneberg, M 2014, ‘Effect of garments on photoanthropometry of body parts: Application of stature estimation’, *Forensic Science International.* Vol. 237, pp. 1-12.

2. Lucas T, Henneberg M 2015a ‘Are human faces unique? A metric approach to finding single individuals without duplicates in large samples’, *Forensic Science International.* DOI: http://dx.doi.org/10.1016/j.forsciint.2015.09.003

3. Lucas T, Henneberg M 2015b ‘Comparing the face to the body, which is better for identification?’, *International Journal of Legal Medicine*. DOI: 10.1007/s00414-015-1158-6.

4. Lucas, T, Henneberg M 2015 ‘Use of units of measurement error in anthropometric comparisons’. Submitted to Journal of Biological and Clinical Anthropology (November 2015).

5. Lucas, T, Kumaratilake, J, Henneberg, M 2015 ‘Metric identification of the same people from images- how reliable is it?’. Submitted to *Journal of Forensic Sciences* (November 2015).

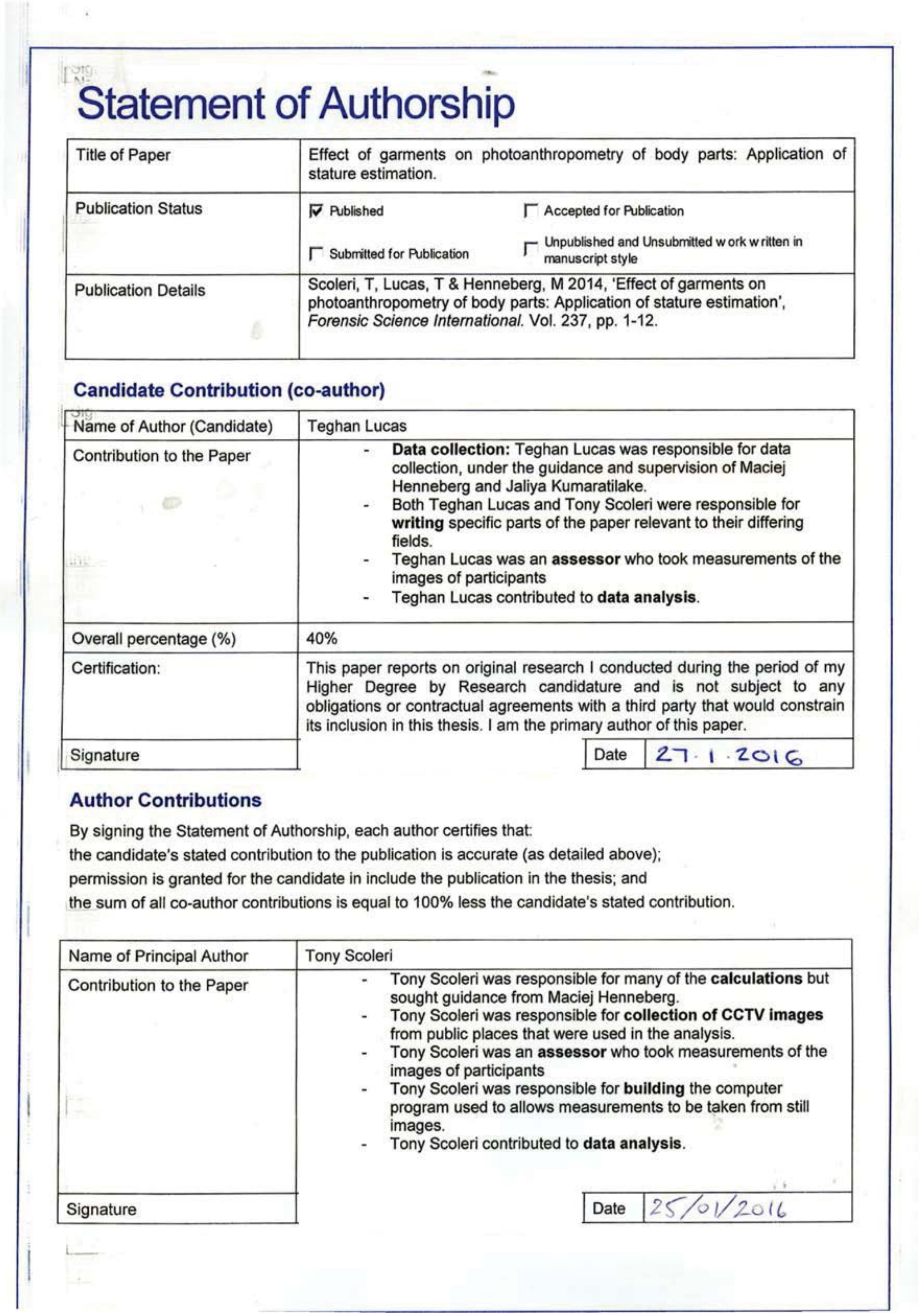
Conference presentations related to this thesis

Lucas, T, Henneberg, M. 2014 ‘Singularity – the absence of duplication without uniqueness’. 28th Annual conference of the Australasian Society for Human Biology in Glenelg, Adelaide.

Note* the title of the paper that this presentation was based on was later changed to ‘Are human faces unique? A metric approach to finding single individuals without duplicates in large samples’ as suggested by reviewers.

Lucas, T, Henneberg, M. 2015 ‘Comparing the face to the body, which is better for identification?’. 84th Annual meeting of American Association of Physical Anthropologists conference in St. Louis, Missouri.

Statement of Authorship Manuscript 1:



Statement of Authorship

Title of Paper	Effect of garments on photoanthropometry of body parts: Application of stature estimation.
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Scoleri, T, Lucas, T & Henneberg, M 2014, 'Effect of garments on photoanthropometry of body parts: Application of stature estimation', *Forensic Science International*. Vol. 237, pp. 1-12.

Candidate Contribution (co-author)

Name of Author (Candidate)	Teghan Lucas
Contribution to the Paper	- **Data collection:** Teghan Lucas was responsible for data collection, under the guidance and supervision of Maciej Henneberg and Jaliya Kumaratilake. - Both Teghan Lucas and Tony Scoleri were responsible for **writing** specific parts of the paper relevant to their differing fields. - Teghan Lucas was an **assessor** who took measurements of the images of participants - Teghan Lucas contributed to **data analysis**.
Overall percentage (%)	40%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 27.1.2016

Author Contributions

By signing the Statement of Authorship, each author certifies that:
the candidate's stated contribution to the publication is accurate (as detailed above);
permission is granted for the candidate in include the publication in the thesis; and
the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Principal Author	Tony Scoleri
Contribution to the Paper	- Tony Scoleri was responsible for many of the **calculations** but sought guidance from Maciej Henneberg. - Tony Scoleri was responsible for **collection of CCTV images** from public places that were used in the analysis. - Tony Scoleri was an **assessor** who took measurements of the images of participants - Tony Scoleri was responsible for **building** the computer program used to allows measurements to be taken from still images. - Tony Scoleri contributed to **data analysis**.
Signature	Date 25/01/2016

Name of Co-Author	Maciej Henneberg		
Contribution to the Paper	- The drafts of the paper were **edited** by Maciej Henneberg. - Added insight into mathematical equations used. - Maciej Henneberg was an **assessor** who took measurements of the images of participants - Maciej Henneberg contributed to **data analysis**.		
Signature		Date	27.1.16

Please cut and paste additional co-author panels here as required.

1. Effects of garments on photoanthropometry of body parts: Application to stature estimation

Tony Scoleri[a], Teghan Lucas[b] and Maciej Henneberg[b].

[a] Image Analysis and Exploitation Group, Defence Science and Technology Organisation, PO Box 1500, Edinburgh 5111, Australia.

[b] Biological and Comparative Anatomy Research Unit, Medical School, University of Adelaide, Adelaide 5005, Australia.

Submitted to Forensic Science International on 30th August 2013, accepted on 30th December 2013, **Published** on 13th January 2014.

Corresponding author:

Tony Scoleri

Image Analysis and Exploitation Group, Defence Science and Technology Organisation, Edinburgh 5111, Australia, Phone: +61 8 7389 7946

Email:tony.scoleri@dsto.defence.gov.au

Context

Surveillance images have been routinely recorded for several decades to maintain security of institutions (i.e. public and private) and public places. They provide evidence for criminal activities carried out by people. To identify a person committing a crime shown on an image, anatomical characteristics of this person must be matched with those of the suspects. Traditionally, this matching has been based on categorical descriptions given by expert witnesses trained in anatomy. Such an approach, could be criticized, since categorical classifications of descriptive traits may be subjective. Therefore, it would be better to replace categorical observations with measurements in interval scales. The problems associated with taking anthropometric measurements from digital images are, technical image distortions, presence of clothing and quality of images. Anthropometry has been used for hundreds of years to identify individuals. Image based evidence is becoming increasingly popular, thus attempts have been made to take measurements from images. Criminals who are aware of the surveillance systems, inevitably cover their faces to avoid identification. Therefore, researchers need to apply anthropometry to other, more visible, parts of the body. It has been suggested that different clothing can alter the perception of an individual’s body shape. However, a study conducted by Lucas, Kumaratilake and Henneberg (2014) found that clothing has very little effect on classification of body shape of males in surveillance images.

Currently, surveillance images are recorded digitally and therefore they can be subjected to computer analysis. The capabilities of computer software in placing anthropometric points on 2D low quality images may produce measurement errors. Education, age and prior experience of persons analyzing images may influence precision with which they will locate anthropometric points and thus influence the accuracy of measurements.

The aims of the research described in this paper were to investigate 1) the effect that different types of garments had on the placement of anatomical points used for anthropometric measurements of the upper limb 2) error rates for the placements of the points.

This research suggested that those working in image analysis should have had specific training in photoanthropometry, which could lead to lower error rates. The results of this study found that thickness and the patterns on the garments produce less accurate results, this would be useful in real life forensic cases.

This method is time consuming and tedious, thus presenting a problem in large scale forensic studies. However, if this method is to be applied for the identification of persons in actual forensic cases, the pool of suspects needs to be narrowed down by other methods initially

Abstract

Person identification from images is an important task in many security applications and forensic investigations. The essence of the problem comes down to measuring key observable anatomical features which can help describing similarities or differences between two or more individuals. In this paper, we examine how different types of garments affect the placement of body markers that enable precise anatomical human description. We focus in particular on landmark positioning errors on the upper limb. Closed-form formulae are provided to compute the maximum likelihood estimate of upper limb length from an image. Subject stature is then predicted from it through a regression model and used as identification criterion. Following initial laboratory experiments, the technique is demonstrated to be invariant to posture and applicable to uninformed subjects in unconstrained environments. Seven technical errors of measurement and statistical tests are quantified empirically from statures obtained by three assessors. Results show that thicker garments produce higher inaccuracies in landmark localisation but errors decrease as placement is repeated, as expected. Overall, comparison to truth reveals that on average statures are predicted with accuracy in excess of 96% for the worse assessor.

Key words: human height estimation, person identification, human characterisation, image measurement, clothing effects, CCTV

Introduction

The use of closed circuit television videos (CCTV) in crime control and prevention has rapidly grown in the present security context. As a counteraction, criminals often disguise their faces to preclude identification. When a person lies within close proximity of a camera, their face features can be acquired with high fidelity. Reliable identification is then achievable, for instance, by finding the smallest distance between the outline of anatomical landmarks on the target face and the corresponding outline of 3-D faces in a gallery (Yoshino et al. 2002). This assumes that the target face is only partially masked and, depending on the viewing geometry and visible facial components, different anatomical points must be identified every time. Further manual intervention is required to register (orientate and scale) the outline of the target face to that of each 3-D face. This task is extremely labour intensive and time consuming the larger the database is.

For the majority of surveillance systems though, the quality of the recorded imagery remains poor. The cost associated with better optical and digital components is often considered too high so many institutions opt for quantity rather than quality (Burton et al.1999). Recorded CCTV images thus display optical distortions, object blur, wrong colours and have low resolution which makes fine details invisible (Chen et al. 2011; Kovesi 2009). To circumvent these issues, attempts have been made using anthropometry to identify individuals based on larger and more easily observable aspects, such as body shape (Henneberg 2007; Henneberg 2008). The use of anthropometric measurements have been included in the description and identification of individuals from photographs as early as the 19th century (Bertillion 1886; Bertillion 1890) and today they are combined with image analytical techniques (Barron and

Kakadiaris 2001; BenAbdelaker and Davis 2006; BenAbdelaker and Yacoob 2008; Cao et al. 2011; Scoleri and Henneberg 2012).

When looking at body shape, the choice of garment has been reported to alter the perception of the human body (Rudd and Chattaraman 2006). In particular, the colour of clothing is noted to have an effect on body appearance. In a survey of male subjects (Frith and Gleeson 2004), it is reported that black is chosen when trying to minimise body size, and light colours, such as white, are worn to maximise body shape definition. The optical illusions presented by patterns such as horizontal stripes are considered to have a contradictory effect by either making the body appear wider and shorter (Johnson 1991) or slimmer and taller (Thompson and Mikellidou 2011). The size of clothing can also affect an individual's perception of body shape, presenting the wearer as larger or smaller than normal depending on the fit of clothing to their actual body parts (Fan and Yu 2004). Latest research using LIDAR technology (McCoppin et al. 2012) has demonstrated that gender classification can be achieved with an accuracy of 70% to 80% for subjects wearing three clothing styles (summer, fall, winter). Another recent study (Lucas, Kumaratilake and Henneberg 2012) has reached similar conclusions, that clothed individuals seen in CCTV images can only be matched above random expectation for a general body shape. These two lines of research have proved independently that the acquisition of body descriptors from images is largely deceived by clothing whether it is obtained from automatic machine learning or based on human perception. Although this statement is intuitive, the challenge is to acquire those descriptors with highest precision from a range of scenarios and body postures.

Our criterion for person identification relies on estimating the body height or stature. When the entire body is visible, image metrology techniques and virtual scene superimposition can be used to measure stature directly from the top of the head to the feet (; Criminisi 1999; DeAngelis et al. 2007; Edelman, Alberink and Hoogeboom 2010). In urban scenes though, pedestrians are often partially occluded in a way that prevents application of such techniques. In other circumstances, only a single image of the person is available rather than a video so stature cannot be derived from gait analysis. Recent research has thus looked at reconstructing the total body height from body parts measurements (Nguyen and Hartley 2012; Scoleri and Henneberg 2012). Clothing style then affects the accuracy to which this may be done. The work in this paper quantifies the effects of garments on photoanthropometry of body parts by first measuring the upper limb length of a person, predicting their stature by linear regression from this length and finally assessing errors from the obtained stature.

In all error functions, part of the error is induced by either the human intervention, when manually locating anthropometric points, or by automatic body part detectors and classifiers. This paper focuses on assessing the human error in the process. Initial laboratory experiments have been conducted on nine male participants wearing no shirt, a black shirt, a horizontally striped shirt and a padded leather jacket. Each of these garments is specifically chosen to evaluate how garment colour, pattern and type alter an assessor’s perception on body landmark location. Subsequent experiments estimate the errors for eighteen uninformed male subjects observed in an uncontrolled (airport) environment. Three assessors have examined the videos and marked points to measure the upper limb length of each subject. The assessors come with varying knowledge in anatomy and computer vision which provides valuable feedback on the spread of errors.

The main contribution of this research is to quantify the variations in placing critical body landmarks when these points are covered by different garment styles (Section 3). Seven error measures are described to estimate those variations and their statistical significance. The second contribution is the overall frame work (Section 2) which includes robust image analytical techniques. In particular, closed-form formulae are proposed for computing the maximum likelihood estimate of the upper limb length (Section 2.2). This framework is advantageous especially for its application to uncontrolled scenes, independence from subject posture and flexibility of integrating other body part measurements as necessary. Contrary to many identification techniques based on facial recognition, the proposed method can be applied to low resolution images and offers limited user interaction (only a few clicks are necessary). This comes as a result of the particular choice of anthropometric landmarks which are distinguishable under arbitrary perspective camera views and suppress the need for intensive image manipulation or alignment. The achievement is that, for stature estimation, average accuracy (or recognition rate) is in excess of 96% for the worst assessor when compared to truth.

Stature estimation

For some years now, researchers in computer vision have tried to find appropriate markers for soft biometry retrieval from videos. Most notable is the work on estimating a person's stature. When the scene is accessible for surveying, a virtual model can be created and superimposed onto the original imagery (DeAngelis et al.2007). As for face recognition, human stature can accurately be measured when subjects walk in vicinity of the camera. In general though, alternative methods are needed in cases where the environment is more challenging (e.g. outdoor), surveying is not possible (e.g. hazardous trafficking area) or subjects are remote from the camera and not standing in perfectly upright position. Besides, whether the person appears in a single frame or is seen in motion through a video sequence, great difficulties arise in precisely extracting the top of the head (vertex) and heel position on the ground

(Aggarwal and Cai 1999; BenAbdelaker, Cutler and Davis 2002; Collins, Gross and Shi 2002; Lee and Choi 2010; Nguyen and Hartley 2012). In all undertakings, the process is image-driven, relying on some variant of silhouette extraction to define the head and feet locations. This resulted in different authors having different definitions for how these points should be retrieved from images. Recent research (Nguyen and Hartley 2012) has settled some of these questions, however, when measuring several partial body dimensions to obtain complete stature, the location of body part markers remains debatable.

In contrast, the scheme we propose is evolved from a well-defined anatomical model of body landmarks extensively tested in anthropological research. The uncertainty in extracting the landmarks from images becomes of prime importance before carrying out any measurement. This is examined in our experiments as well as its effect on a person's stature estimation. Forensic scientists and anthropologists routinely perform measurements of long bones to reconstruct total body height from regression equations that relate those body parts to the human stature (Blau and Ubelaker 2009; Byers and Myster 2009) Such reconstructions have been widely successful and achieve about 95% accuracy on body height prediction (Willey 2009). Statistically, the upper limb length relates significantly to the human stature and is commonly observable and measurable in CCTV videos. Besides, its length does not undergo any diurnal change unlike stature. It is indeed a well-recognised phenomenon that stature begins to decrease immediately after rising in the morning and further loss continues throughout the day up to a maximum of 28.1mm (Krishan and Vij 2007). A person should be measured preferably in the afternoon to reduce the variation in stature as loss of height occurs most rapidly in the morning (Krishan and Vij 2007; Voss and Bailey 1997). Since this constraint is not realisable in general surveillance context, the upper limb length presents a good alternative to infer stature.

Upper limb model

The upper limb length of a subject is measured as

$$u = \mathrm{sgn}(h_a - h_d).(h_a - h_d),$$

where sgn(x) stands for the signum function of a real number x, ha is the height from a ground point to the acromiale (the shoulder point) and hd is the height from the same ground point to the dactylion (the tip of the middle finger). These heights are obtained from an image by asking a user to place three markers corresponding to the acromiale, dactylion and a point on the ground. The use of the signum function enables to mark the acromiale and dactylion in any order of preference. One assumption here is that these two points are situated at the same depth in the scene and therefore define an upright vertical segment to the ground. Their depth is estimated as the distance from the third point placed on the ground and a world origin[11]. Once the acromiale and dactylion points are set, their corresponding Maximum Likelihood (ML) location is computed such that the ML points are aligned with the vertical scene direction, see Figure 1. Details of the alignment procedure are deferred to the next section. To assist the user in choosing the ground point, a guide line is drawn through the ML points. The ground marker can then be set along the line where the heels of the subject are touching the floor. A short procedure is also applied to ensure that this point lies on the guide line perfectly. With the three collinear points, heights h_a and h_d are orthogonal to the ground and can be calculated as described in Section 2.3.

---

[1] The world origin is chosen on the ground during camera calibration.

Figure 1. Upper limb model. The user-marked points are shown in red and computed ML points in green.

Maximum Likelihood estimation of limb endpoints

Image perspective, subject posture and user interaction mean that the upper limb segment will rarely stand in the vertical scene direction required to measure its length. This limitation is addressed by computing new endpoints which are aligned with the vertical vanishing point, $\mathbf{v}_z$, obtained during calibration.

Suppose that the input markers are given by two points $\mathbf{x} = [x_1, x_2]^{\mathrm{T}}$ and $\mathbf{x}' = [x_1', x_2']^{\mathrm{T}}$ with associated 2 x 2 isotropic Cartesian covariance matrices $\mathbf{\Lambda}_{\mathbf{x}}$ and $\mathbf{\Lambda}_{\mathbf{x}'}$ defining perturbation around $\mathbf{x}$ and $\mathbf{x}'$ by circles of radius r and r', respectively. The Maximum Likelihood estimates $\bar{\mathbf{x}}$ and $\bar{\mathbf{x}}'$ of input markers $\mathbf{x}$ and $\mathbf{x}'$ can be determined by minimising the squared Mahalanobis distance

$$d_{\mathrm{Mahal}}^2 = (\mathbf{x} - \bar{\mathbf{x}})^{\mathrm{T}} \Lambda_{\mathbf{x}}^{-\mathbf{1}} (\mathbf{x} - \bar{\mathbf{x}}) + (\mathbf{x}' - \bar{\mathbf{x}}')^{\mathrm{T}} \Lambda_{\mathbf{x}'}^{-\mathbf{1}} (\mathbf{x}' - \bar{\mathbf{x}}')$$

subject to the alignment constraint $\mathbf{v}_z^T \boldsymbol{l} = 0$, with $\boldsymbol{l} = [\bar{\mathbf{x}}^{\mathrm{T}}, 1]^{\mathrm{T}} \times [\bar{\mathrm{x}}'^{\mathrm{T}}, 1]$. This is a constrained optimization problem which can be solved in closed-form using the Lagrange multiplier method. For $\mathbf{v}_z = [v_1, v_2, v_3]^{\mathrm{T}}$ and $\mathbf{z} = [r, r', \mathbf{x}^{\mathrm{T}}, \mathbf{x}'^{\mathrm{T}}, \mathbf{v}_z^T]^{\mathrm{T}}$, it can be shown that

$$\boldsymbol{l} = \begin{bmatrix} 1 + \sqrt{1 + [\xi(\mathbf{z})]^2} \\ \xi \\ -v_1 v_3^{-1} \left(1 + \sqrt{[\xi(\mathbf{z})]^2}\right) - v_2 v_3^{-1} \xi(\mathbf{z}) \end{bmatrix},$$

where the real-valued rational function $\xi: \mathbb{R}^9 \mapsto \mathbb{R}$ has the form

$$\xi(\mathbf{z}) = 2 \frac{r' d_1 d_2 + r d_1' d_2'}{r'(d_1^2 - d_2^2) + r(d_1'^2 - d_2'^2)}$$

with $d_i = x_i - v_i v_3^{-1}$, $d_i' = x_i' - v_i v_3^{-1}$, $i = 1, 2$. The previous formulae hold as long as the vertical vanishing point is not ideal ($v_3 \neq 0$).

In our implementation, anisotropic anisotropic Cartesian covariances $\widetilde{\mathbf{\Lambda}}_{\mathbf{x}}$ and $\widetilde{\mathbf{\Lambda}}_{\mathbf{x}'}$ are first calculated (Brooks et al. 2001) and then employed to yield

$$r = \left|\det(\widetilde{\mathbf{\Lambda}}_{\mathbf{x}})\right|^{1/4} \quad r' = \left|\det(\widetilde{\mathbf{\Lambda}}_{\mathbf{x}'})\right|^{1/4}.$$

Writing $\boldsymbol{l} = \left[l_x, l_y, l_w\right]^{\mathrm{T}}$, the ML estimates of $\mathbf{x}$ and $\mathbf{x}'$ are given by the Cartesian coordinates

$$\bar{\mathbf{x}} = \left[\frac{x_1 l_y^2 - x_2 l_x l_y - l_x l_w}{l_x^2 + l_y^2}, \frac{x_2 l_x^2 - x_1 l_x l_y - l_y l_w}{l_x^2 + l_y^2}\right]^{\mathrm{T}}$$

$$\bar{\mathbf{x}}' = \left[\frac{x_1' l_y^2 - x_2' l_x l_y - l_x l_w}{l_x^2 + l_y^2}, \frac{x_2' l_x^2 - x_1' l_x l_y - l_y l_w}{l_x^2 + l_y^2}\right]^{\mathrm{T}}$$

These points are taken as the true locations of the upper limb endpoints. The above derivation differs from its original form in Criminisi (1999) in that it critical entities are readily programmable as given here with some of them explicitly calculated in Cartesian rather than projective coordinated to prevent potential errors.

Upper limb length

Without loss of generality, suppose we wish to determine the actual height h_a from the ground point $\boldsymbol{G}$ to the acromiale $\boldsymbol{A}$. Let $\boldsymbol{g}$ and $\boldsymbol{a}$ denote their corresponding image points, with $\boldsymbol{a}$ the ML estimate of the user-defined acromiale. Assuming a perspective projection camera model, these relationships may be written as

$$\{\lambda_1[\boldsymbol{a}^T, 1]^T = \mathrm{P}[\boldsymbol{A}^T, 1]^T, \lambda_2[\boldsymbol{g}^T, 1]^T = \mathrm{P}[\boldsymbol{G}^T, 1]^T$$

where P encodes the projection matrix and the λ_i's some perspective scale factors. The above system of equations can be expressed in a matrix form as M$\boldsymbol{A}$=$\mathbf{b}$. This follows from using the assumption that $\mathbf{A}$ and $\mathbf{G}$ are at the same depth, so one may write $\mathbf{A} = [\mathrm{A}_1, \mathrm{A}_2, \mathrm{h_a}]^{\mathrm{T}}$ and $\mathbf{G} = [\mathrm{A}_1, \mathrm{A}_2, 0]^{\mathrm{T}}$ which provides four equations in three unknowns.

The least-squares solution $\hat{\boldsymbol{A}}=\left[\mathrm{M}^T\mathrm{M}\right]^{-1}\mathrm{M}^T\mathbf{b}$ gives h_a as the third component of $\hat{\boldsymbol{A}}$. Height h_d can be calculated in a similar manner. The upper limb length ensues from the formula given in Section 2.1.

Anthropometric stature prediction

Anthropometric data were collected from 109 adult males resident in Australia. These included upper limb length and body height measured in accordance with the Martin's Technique (Martin and Saller 1957) and in compliance with the International Standards Organisation (ISO 7250). Two linear regression techniques, namely the Ordinary Least-Squares (OLS) and Reduced Major Axis[2] (RMA), can be applied to these data to predict stature from upper limb length. If u and s denote the upper limb length and stature respectively, both expressed in millimetres (mm), then

$$\text{OLS:}\quad s = 1.4052u + 678.74,$$

$$\text{RMA:}\quad s = 1.7435u + 413.94.$$

The variances associated with the OLS and RMA stature predictions are 48.5 mm and 50.6 mm, respectively.

Experiments

Clothing effects are first examined under controlled conditions in a laboratory (Sections 3.1 to 3.3). Various qualitative measures are calculated to evaluate the errors in landmark placement and differences between assessors. Visual influence of garment

---

[2] RMA is also known as the Total Least-Squares method.

styles is deduced from those errors. Section 3.4 reports the results when our technique is applied to uninformed subjects in an unconstrained (airport) scene. In all laboratory tests, only the OLS regressor is used to predict stature from upper limb length. The RMA method is temporarily discarded because RMA statures are linearly related to OLS ones and therefore would reveal similar trends. RMA is used in Section 3.4 on every-day life surveillance imagery taken in an airport.

Laboratory experiments overview

Laboratory set-up

Nine adult male participants have been recruited within South Australia. Each of them is made to wear a pair of surgical pants, shoe coverings, a cap and a face mask to eliminate identifying features. This strategy also intends to focus the assessors' attention on the upper body with no other distractions. Participants are recorded using a CCTV camera (axis p3304 with a resolution of 1280 x 800 pixels) in the Bioskills laboratory of the Medical School at The University of Adelaide. The camera is fixed to the ceiling at a height of 2.5 metres from the floor. Still photographs are extracted from the videos and calibrated using the technique in (Scoleri 2010). Each participant is standing approximately 8 metres from the camera shown in either an anterior or posterior view. They are imaged four times wearing no shirt, a black shirt, a horizontally striped shirt and a padded leather jacket (n = 36 photographs). Figure 2 shows the garments on a participant. In order to reduce the influence of diurnal variation on stature, the men have been measured and recorded in the afternoon. This provided an estimate of their "true" stature. The quotation marks are used because although every care has been taken to minimise the errors in true statures, it was reported that some may still exist (Ulijaszek and Lourie 1994).

Figure 2. Anterior views of the same male in four different types of wear. (a) Shirtless; (b) Black shirt; (c) Striped shirt; (d) Padded jacket

Several challenges became apparent when analysing the images. Despite instructing the participants on the correct pose to hold, they are often slouching or leaning to one side, their upper limb is slightly bent and not straightened out, their hands are curled up and not fully opened, their feet are not together but separated (Figure 3). These various postures do not adhere to the correct anatomical model and therefore introduce errors. However, they offer realistic conditions as would be encountered in real-life CCTV images, which is important.

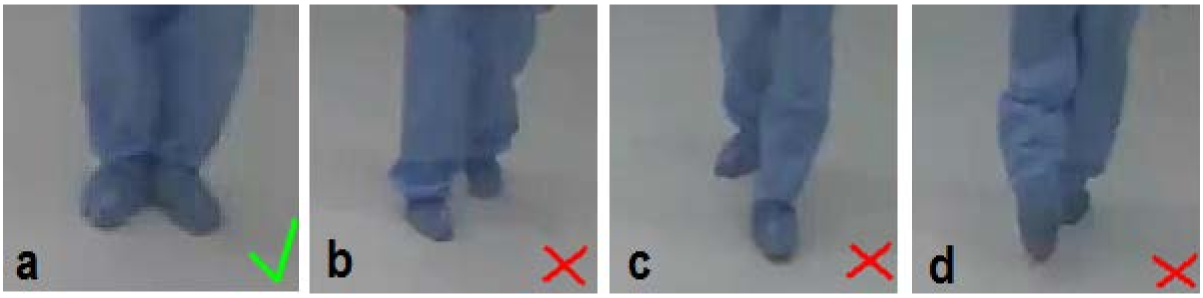


Figure 3. Sample images of feet position for different participants. (a) Anatomically correct position; (c-d) Incorrect pose.

Assessors

Three assessors have viewed each of the photographs and marked the three points as described in Section 2.1. Their level of experience in anatomy can be ranked as expert, trained and novice. The “expert” has over forty years of experience studying and measuring the human body and is employed as a Professor of Anatomical Sciences, teaching students as well as conducting research in the field. The person who is referred to as “trained” is educated in the field of Biological Anthropology and has three years of experience in measuring and studying the human body. The “novice” assessor, although expert in computer vision gait analysis, has limited experience in defining a person’s anatomical features. The broad range of expertise is valuable for observing the variations of errors. In the experiments, assessors only have a single attempt at marking points for each participant. This guarantees integrity in revealing the effects of garments on landmark placement. It also means that error values can improve if an assessor could mark points repeatedly. Assessor 2 was chosen to complete the task twice in order to calculate the intra-observer error. The landmark positioning took just over an hour for each assessor to go through the 36 images; assessor 2 was given a two-hour break in between repeats in order to reduce the effects of fatigue and memory on the task.

Error analysis for three uncertain markers

The Technical Error of Measurement (TEM) is a suitable quality measure to assess the difference between stature measurements (Ulijaszek and Lourie 1994). Several such TEMs are described in the following sections. In addition, the bias between assessors’ measurements and the significance of variances between TEM values are examined. In total, seven tests are presented to understand how errors fluctuate under the influence of landmark positioning and garment style.

Comparison to truth

The first TEM evaluates the discrepancy between all predicted statures and their “truth” values. It is given by the formula:

$$\mathrm{TEM}_{\mathrm{truth}} = \left(\frac{1}{2n} \sum_{i=1}^{n} (\bar{s}_i - s_i)^2 \right)^{1/2},$$

where n is the total number of test images, $\bar{s}_i$ and s_i stand for the true and predicted statures, respectively. The assessors produced TEMs as shown in Table 1.

Table 1. Errors (mm) between predicted statures and truth.

	Assessor 1	Assessor 2	Assessor 3
$\mathrm{TEM}_{\mathrm{truth}}$	30.0	39.2	44.3

In our model, the total error on stature prediction stems from a combination of the errors in placing the anthropometric points, calibrating the camera and the OLS regressor. Given that the variance for OLS is estimated at 48:5 mm, these errors are within that threshold and thus very encouraging. Considering the worst score and comparing to the average participant's height of 1758 mm yields an accuracy in excess of 97%. For the best TEM score, the accuracy reaches over 98%.

Inter-observer error

An inter-TEM is used to measure the extent to which predictions from assessors differ from one another:

$$\mathrm{TEM}_{\mathrm{inter}} = \left(\frac{1}{n(k-1)} \sum_{i=1}^{n} \left[\sum_{j=1}^{k} s_{i,j}^2 - \frac{\left(\sum_{j=1}^{k} s_{i,j}\right)^2}{k} \right] \right)^{1/2},$$

where s_{ij} is the *i*-th predicted stature obtained by the *j*-th assessor and *k* refers to the total number of assessors (here *k* = 3). Table 2 shows the values obtained for different selections of assessors. More experienced anatomists (assessors 1-2) have a lower TEM than less experienced anatomists (assessors 2-3). So, these results confirm the assessor's experience in anatomy.

Table 2. inter-TEM (mm) for selections of assessors

	Assessors			
	1 – 2	1 – 3	2 – 3	1 – 2 – 3
$\text{TEM}_{\text{inter}}$	23.5	26.5	27.4	25.9

Intra-observer error

Assessor 2 has performed the point marking twice for all nine participants. This allows the intra-TEM to be measured as

$$\text{TEM}_{\text{intra}} = \left(\frac{1}{2n} \sum_{i=1}^{n} (s_i^1 - s_i^2)^2 \right)^{1/2}$$

where s_i^1 and s_i^2 are the predicted statures obtained in the first and second round of marking, respectively. The intra-TEM may be interpreted as an indicator of the measurements' reliability. As measurements are repeated, some variations are initially expected until a point where the error is reduced to a small value and progress can no longer occur. When the intra-TEM stagnates, one can be confident about the predicted statures. For assessor 2, the $\text{TEM}_{\text{intra}}$ is found to be equal to 35. 1 mm. This value suggests that possible improvement of the landmark positioning can be made.

Assessor's bias

The bias between two assessors placing markers can be quantified as

$$\text{bias}_{a-b} = \frac{1}{n} \sum_{i=1}^{n} \left(s_i^a - s_i^b \right)$$

where s_i^a and s_i^b are the *i*-th predicted statures obtained from assessor a and b, respectively. The variables *a* and *b* take distinct values in the range 1,...,*k*. Results are summarised in Table 3. Again, the more experienced anatomists (pair 1-2) recorded much smaller bias than less experienced ones (pair 1-3). Looking at pair 1-2, the positive value for the bias means that on average the first assessor predicted taller statures than the second one. A similar reasoning can be deduced regarding the other two pairs.

Table 3. Bias (mm) between different pairs of assessors.

	Assessors		
	1 – 2	1 – 3	2 – 3
Bias	+0.1	+22.2	+22.1

Effects of garments

TEM$_{truth}$ provides an error measure which is too generic and does not reveal the effect of a particular garment on the assessors' ability to mark the required points. The analysis in this section addresses this limitation. First, the inter-TEM is calculated by including the measurements that only relate to a specific clothing style: (a) All participants are shirtless, (b) with a black shirt, (c) a striped shirt or (d) a padded jacket (n = 9). Table 4 presents the results. Overall, the TEM for shirtless participants turn out to be the largest due to the roundness of the shoulder and thus the increased ambiguity to mark the acromiale. Lowest TEM is achieved for participants wearing a black shirt as it defines the silhouette better around the shoulders. When looking at the various garment types, those with stripes or padded produce higher inaccuracies, which is to be expected.

Table 4. Inter-TEM (mm) for (a) shirtless participants; (b) with black shirt; (c) with striped shirt; (d) with padded jacket.

	Assessors			
	1 – 2	1 – 3	2 – 3	1 – 2 – 3
(a)	30.0	28.3	37.7	32.3
(b)	16.6	26.0	10.0	18.7
(c)	22.5	21.3	25.8	23.3
(d)	22.8	29.8	28.7	27.3

In a second series of tests, the bias is examined by comparing the shirtless case to the clothed ones. The formula in Section 3.2.4 is used with s_i^a taken as the stature measurement obtained for a shirtless participant and s_i^b as the measurement for the corresponding participant wearing either the black shirt, the striped shirt or the padded jacket for all assessors (n = 27). Results are given in Table 5.

Table 5. Bias (mm) between different garment types. The abbreviation 'SL' stand for 'Shirtless'.

	SL-Black	SL-Striped	SL-Padded
Bias	+14.5	+36.4	+37.8

As can be seen, the difference in marking anthropometric points is significantly smaller when shirtless participants are compared to those wearing a black shirt. The striped shirt and padded jacket increase the difficulty in identifying points which yields larger errors in both cases. The consistently positive bias across all three categories indicates that the estimated statures are taller on average for shirtless participants and thus may suggest a tendency to place markers more incorrectly in this situation. This would agree with the results in Table 4 where the inter-TEMs (almost) always show greater variation in the shirtless case than in the other three cases.

Snedecor's F-test

Technical errors of measurements are essentially variances of one measurement around another measurement. Their random errors are thus a result of the measurement-to-measurement differences and sample sizes. This means that the difference between two TEM values can be tested for statistical significance in the same way the difference of two variances is. The Snedecor's F-test is an appropriate tool to use. This test is based on the ratio of two variances in general populations to assess significance because the distribution of errors of ratios of larger to smaller variances depends on the combination of their degrees of freedom that are determined by sample sizes minus one. The F-test is given by

$$\mathrm{F} = \frac{v_1}{v_2}$$

where v_1 is the larger variance, v_2 the smaller variance. Both v_1 and v_2 are estimates of population variances derived from sample values in the following way:

$$v_1 = \frac{v_1'.N}{N-1}, \qquad v_2 = \frac{v_2'.N}{N-1}$$

with v_1' and v_2' the sample variances and N the number of observations. TEMs are square roots of sample variances, hence after squaring TEM values, multiplying them by N and dividing the result by N - 1, we obtain equivalents of variances appropriate to form ratios for the F-test. Note that with large numbers of observations, the $N/(N - 1)$ term approaches 1 and thus direct ratio of squared TEMs is an approximation of the F-test value. For the TEMs discussed in Sections 3.2.1, 3.2.2 and 3.2.5, with the number of observations $N = 9$, most squared TEM ratios do not exceed appropriate cut-off F-test values for the 0.05 significance of differences. In Table 4, the TEMs in rows (a)-(b) for assessors 1-2 and 2-3 are the two instances where the F-tests are statistically different. This is because locating landmarks is much easier when a person wears a black shirt than no shirt, as explained in Section 3.2.5.

Error analyses for a single marker

Among the three markers to place, only one of them is truly covered by clothing: the acromiale. In this section, we investigate the errors and effects of garments on this particular landmark. Since all three markers are recorded per assessor for all photographed subjects, we conducted a first series of tests whereby the dactylion and ground points are always taken as those from the *expert* assessor. This means that the location of the acromiale remains as chosen by the individual assessor. The error measures presented in Section 3.2 are labelled with a superscript $^{\text{exp}}$ to mark this distinction. We have also recalculated all the errors when the dactylion and ground points originate from the *novice* assessor. These are labelled with a superscript $^{\text{nov}}$. The two data manipulation strategies are employed to examine how the errors fluctuate for radically different expertise levels and whether it reveals any pattern.

Comparison to truth

Table 6 summarises the results for $\text{TEM}_{\text{truth}}$. The values in brackets indicate the relative difference with the results in Table 1 when all three markers are chosen by each assessor. Since Assessor 1 is the person with expert anatomical knowledge, his score for $\text{TEM}_{\text{truth}}^{\text{exp}}$ remains unchanged from Table 1 (identical data). A considerable improvement can be noted for the novice anatomist (Assessor 3) who progresses to a comparable level to that of the expert assessor. The TEM for the trained anatomist decreased minimally. This overall trend is to be expected since Assessor 1 has a better selection of points (smallest error to truth in Table 1).

Table 6. $\text{TEM}_{\text{truth}}$ (mm) when then dactylion and ground points are those from the expert and novice assessors.

	Assessor 1	Assessor 2	Assessor 3
$\text{TEM}_{\text{truth}}^{\text{exp}}$	30.0 (+0.0)	38.7 (-0.5)	32.9 (-11.4)
$\text{TEM}_{\text{truth}}^{\text{nov}}$	44.1 (+14.1)	48.9 (+9.7)	44.3 (+0.0)

In the second row, the fixed dactylion and ground points come from Assessor 3 so his score is unchanged from Table 1. The increase in value for the results of the other assessors is understandable given that Assessor 3 has the largest discrepancy to truth when all three markers are specified. Looking globally at the results, the relative difference between assessors 1 and 3 is negligible for both $\mathrm{TEM}_{\mathrm{truth}}^{\mathrm{exp}}$ and $\mathrm{TEM}_{\mathrm{truth}}^{\mathrm{nov}}$. This suggests that they have consistently placed the acromiale around the same location. The difference in landmark placement is more noticeable for assessor 2 who has a larger residual error in both tests.

Inter-observer error

Table 7 shows the new inter-TEM values. The numbers in brackets indicate the difference with the results in Table 2 when all three markers are chosen freely by each assessor.

Table 7. Inter-TEM (mm) for selections of assessors.

	Assessors			
	1 – 2	1 – 3	2 – 3	1 – 2 – 3
$\mathrm{TEM}_{\mathrm{inter}}^{\mathrm{exp}}$	18.6 (- 4.9)	12.2 (- 14.3)	17.5(- 9.9)	16.3 (-9.6)
$\mathrm{TEM}_{\mathrm{inter}}^{\mathrm{nov}}$	18.7 (- 4.8)	12.6 (-13.9)	17.1 (-10.3)	16.4 (- 9.5)

All errors have decreased and turned out about the same magnitude. The smaller variations between assessors are a direct consequence of fixing two of the three markers. Assessors 1-3 produced the smallest inter-TEM values whereas the largest values are observable when statures from assessor 2 are compared to those of assessors 1 and 3. Since the only source of uncertainty arises from the location of the acromiale, these results confirm that assessors 1 and 3 placed similar landmarks, and that assessor 2 was visually more affected by the clothing styles.

Assessors' bias

The bias is now calculated for the new data, refer to Table 8. All errors are small and about the same magnitude. Statures from assessor 3 are shorter than those of assessor 1 (positive bias) but taller than those of assessor 2 (negative bias) with similar amount of variation in each case. From column 1, assessor 2 yielded shortest statures. This agrees with prior findings that this assessor has tangibly different point locations than the other assessors. Assuming equal difficulty in marking the acromiale and dactylion, one may deduce from the bias values in Tables 3 and 8 that a large part of the error stems from the location of the ground point. Assessor 2 must have placed this point much better than assessor 3 because all relevant errors in Section 3.2 are larger for assessor 3 and we have identified in this section that his placement of the acromiale is comparable to the expert anatomist.

Table 8. Bias (mm) between different pairs of assessors.

	Assessors		
	1 – 2	1 – 3	2 – 3
Bias^{exp}	+3.8	+2.0	-1.8
Bias^{nov}	+4.0	+1.8	-2.2

Effects of garments

Following the analysis in Section 3.2.5, we consider the inter-TEMs for statures that only relate to a particular type of clothing. Results are summarised in Tables 9 and 10.

Table 9. $\mathrm{TEM}_{\mathrm{inter}}^{\mathrm{exp}}$ (mm) for (a) shirtless participants; (b) with black shirt; (c) with striped shirt; (d) with padded jacket.

	Assessors			
	1 – 2	1 – 3	2 – 3	1 – 2 – 3
(a)	26.6	14.2	16.2	19.8
(b)	13.4	15.8	13.3	14.2
(c)	13.0	8.7	18.5	14.0
(d)	17.9	8.3	21.0	16.7

Table 10. $\mathrm{TEM}_{\mathrm{inter}}^{\mathrm{nov}}$ (mm) for (a) shirtless participants; (b) with black shirt; (c) with striped shirt; (d) with padded jacket.

	Assessors			
	1 – 2	1 – 3	2 – 3	1 – 2 – 3
(a)	27.0	14.2	16.0	19.9
(b)	14.1	16.5	13.6	14.8
(c)	12.2	8.7	17.3	13.2
(d)	17.9	9.1	20.9	16.7

Performing a row-wise comparison between the two tables, we see that all errors have about the same order of magnitude and have decreased compared to their corresponding values in Table 4. In relation to clothing effects, the decrease is more significant for striped shirt and padded jacket which demonstrates that these garments present a greater challenge for someone to place markers correctly. This corroborates the conclusion in Section 3.2.5. Other experiments focus specifically on the errors between different clothing styles. Looking at Table 11, the effect of a particular garment is clearly visible since most often the bias is positive with large magnitude. The trends are shown more prominently here compared to Table 5 because the uncertainty is assessed precisely in the marker covered by clothing. Note that the bias in column 1, row 2, is small and negative. This peculiarity is a consequence of using the ground points from assessor 3, which we know from previous experiments are not well placed. This result may be ignored.

Table 11. Bias (mm) between different garment types. The abbreviations ‘SL’ stands for ‘shirtless.

	SL-Black	SL-Striped	SL-Padded
Bias^{exp}	+33.4	+45.1	+34.7
Bias^{nov}	-5.3	+31.8	+38.0

Snedecor’s F-test

F-tests have been performed in a similar fashion to those in Section 3.2.6. Looking at Tables 2 and 7, the TEMs for the assessors’ pair 1-3 exhibit statistically significant differences at the 95% confidence level. This result is a formal confirmation that the provision of better ground points has improved the stature estimates of assessor 3 and thus reduced the inter-TEM with assessor 1 noticeably. Other statistically significant differences can be found in both Tables 9 and 10 between rows (a)-(b) and (a)-(c) for assessors’ pair 1-2, between rows (b)-(d) for pair 1-3 in Table 9 and rows (b)-(c) for pair 1-3 in Tables 10. Since acromiale is the only variable point, the F-tests prove that assessors 1-2 have consistently placed that point when participants wear shirts (low inter-TEMs in rows (b), (c)) compared to no shirt (row (a)). For assessors 1-3, the F-tests reveal that they had similar point marking for participants wearing a striped shirt and a padded jacket (rows (c), (d)) compared to when they wear a black shirt (row (b)).

Real-life surveillance images

Our model is now applied to an airport surveillance video released for the Performance Evaluation of Tracking and Surveillance 2007 workshop (PETS 2007). We have examined 4,500 images (of resolution 720 x 576 pixels) of a particular video clip. Over 95% of passengers and bypassers are found to be missing lower limbs due to luggage and other people obstructing. To have benchmark statures for comparison, we have considered the same 30 men as those in (Scoleri and Henneberg 2012), however several challenges appeared immediately:

1. Some bypassers wear clothes of colour similar to the scene background. This especially precludes the marking of the acromiale;
2. Many waiting passengers stand with their arms crossed over their chest or behind their back, their upper limb is partially occluded, their hands are closed or hidden;
3. Image resolution is too poor to clearly distinguish body parts of pedestrians far in the scene. Markers would need to be placed with sub-pixel accuracy, which is not straightforward to do.

In these situations, the point placement is not trustworthy or possible, so the men are discarded. Such situations are illustrated in Figure 4. This brings the number of test subjects down to 18. Two of these subjects are viewed facing the camera, one from the back and fifteen others under various side-way postures. The latter postures are most difficult to deal with, even when the complete upper limb is visible (Figure 5). It is indeed easier to locate the acromiale when both shoulders are observable as in our laboratory experiments where participants are in anterior or posterior view. Among the test subjects, four passengers have garments of colour similar to the background and two others walk with their arms bent pushing trolleys. We kept these six candidates to see how the assessors and our model would cope with extreme situations. Figures 5 and 6 depict some workable examples and other more challenging cases in our test set.

In order to carefully examine the effects of garments, the 18 selected men are separated into six categories of equal size. In this context, garment style but also subject posture must be taken into account, the latter being a novel addition compared to the laboratory experiments. People for whom the upper limb is reasonably visible are considered as having an adequate posture; otherwise, they are labelled as having inadequate posture. The categories are:

(a) *Easy-shirt:* people are wearing a shirt and stand with adequate posture;

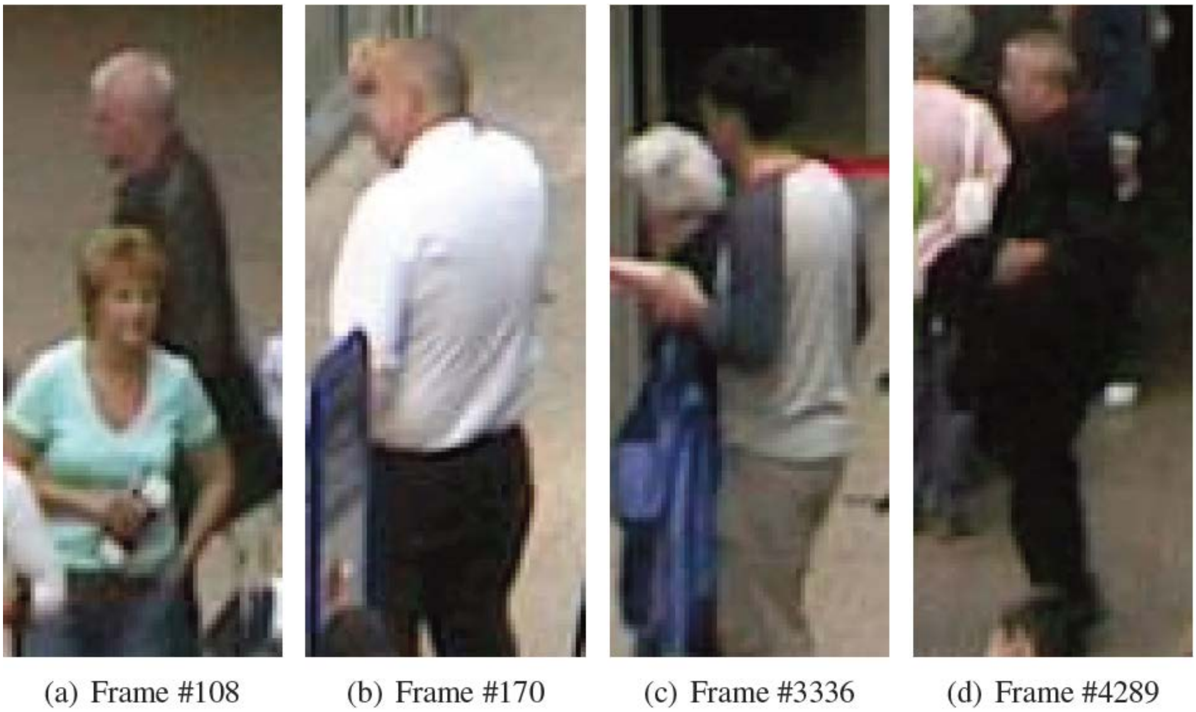
(b) *Easy-jumper:* people are wearing a jumper and stand with adequate posture;

(c) *Easy-jacket:* people are wearing a padded jacket or thick coat and stand with adequate posture;

(d) *Hard-shirt:* people are wearing a shirt and stand with inadequate posture;

(e) *Hard-jumper*: people are wearing a jumper and stand with inadequate posture;

(f) *Hard-jacket*: people are wearing a padded jacket or thick coat and stand with inadequate posture;



(a) Frame #108 (b) Frame #170 (c) Frame #3336 (d) Frame #4289

Figure 4. examples of subjects discarded from our experiments.

(a) Frame #1458 (b) Frame #2766 (c) Frame #3779 (d) Frame #4260

Figure 5. Subjects with visible upper limb.

(a) Frame #1623 (b) Frame #2786 (c) Frame #3023 (d) Frame #4044

Figure 6. Challenging situations for point marking.

Experimental arrangements

In the test images, all 18 passengers have partially occluded legs and feet, so no ground point is markable. The authors in (Scoleri and Henneberg 2012) have kindly provided the camera calibration, a list of subjects, their predicted statures from head heights, top and base head point locations and their projections onto a reference plane. Our experimental set-up could thus replicate their exact conditions. We show next how the ground point entering the calculation of the upper limb length is obtained from this starting information.

The world coordinate system is set such that the X-Y plane is on the ground and the positive Z-axis represents the upward vertical scene direction. Let $\mathrm{o_w}$ be the image of the world origin, **t** the head top point and $\tilde{\mathbf{t}}$ its projection onto a reference plane perpendicular to the ground plane (Figure 7). Furthermore, let $\mathrm{v_x}$, $\mathrm{v_y}$, $\mathrm{v_z}$ denote the vanishing points in the X-, Y-, Z-directions, respectively. Viewing the 3-D world as a collection of three orthogonal pencils of parallel planes (Criminisci1999), it can be shown that the projection of $\tilde{\mathbf{t}}$ onto the ground plane, which is aligned with **t** and $\boldsymbol{v}_z$, is the homogeneous point

$$\mathrm{g} = ([\mathrm{t}^{\mathrm{T}}, 1]^{\mathrm{T}} \times \mathrm{v_z})\ \mathrm{x}\ (\mathrm{v_x} \times \mathbf{m})$$

with $\mathbf{m} = \left([o_w^T, 1]^T \times v_y\right) \times ([\tilde{\mathbf{t}}^T, 1]^T \times v_z)$. Point **g** is then projected orthogonally onto the guide line formed by the (ML) acromiale and dactylion points (Figure 7). This technique yields a valid ground point for upper limb measurement. Only two points are now required, the acromiale and the dactylion. As in the laboratory experiments, assessors have marked each point in a single action. This is a major improvement over methods which require extensive repeats of the point placement and need to operate at a sub-pixel level (Scoleri and Henneberg 2012).

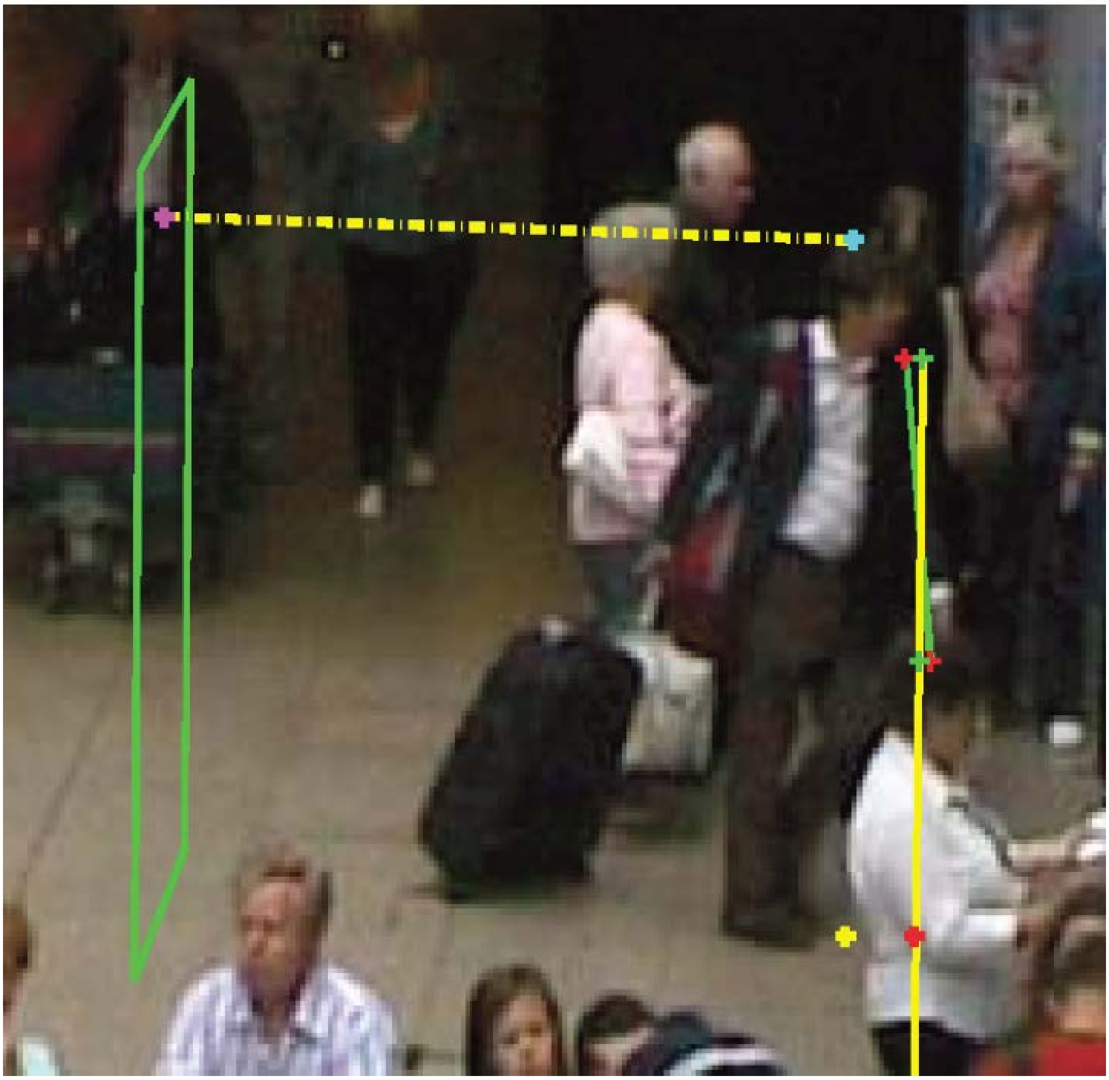


Figure 7. Construct of upper limb length measurement with point $\tilde{\mathbf{t}}$ (magenta) on the reference plane (green), **t** (cyan) on the head vertex, **g** (yellow) on the ground and its projection (red) onto the guide line.

Test results

The same three assessors as those in Section 3.1 have placed markers to obtain the upper limb lengths for all subjects. These lengths are subsequently used to infer statures through OLS and RMA regressions (Section 2.4). The final statures are taken as the average values of the two predictions to match the approach in (Scoleri and Henneberg 2012) and compare estimates. Figures 8 and 9(a) show the results for all three assessors along with the predicted statures from (Scoleri and Henneberg 2012).

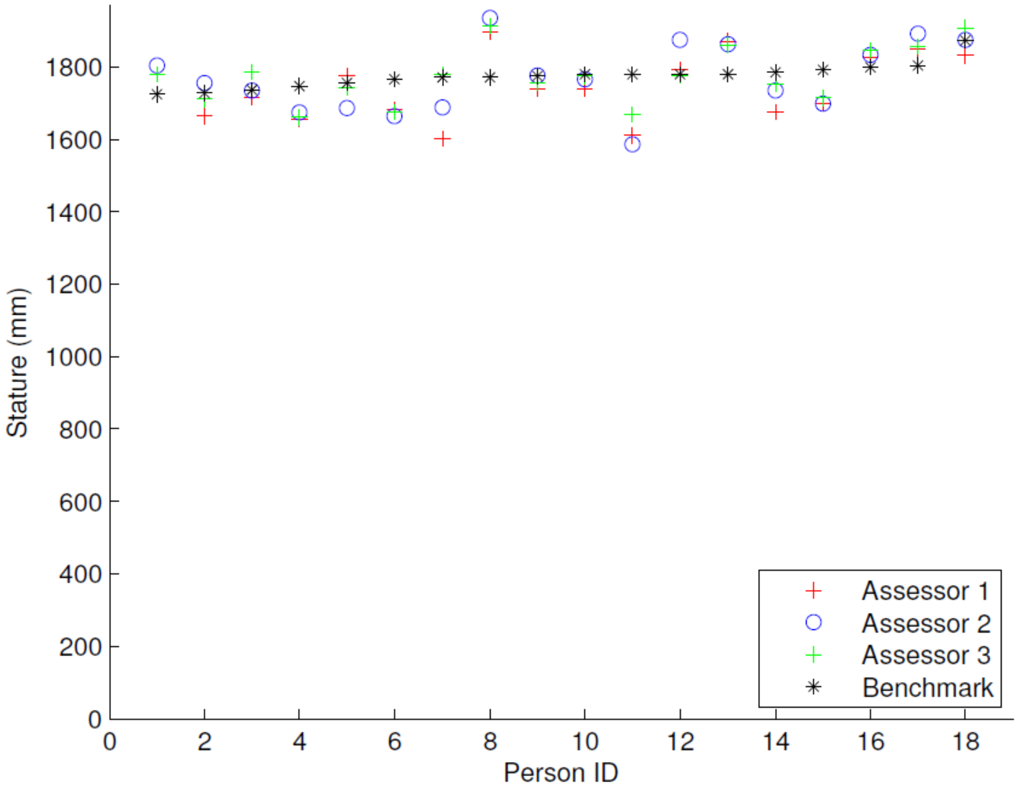


Figure 8. Stature estimates for 18 passengers. Benchmarks statures originate from (Scoleri and Henneberg 2012).

One may consider the statures from (Scoleri and Henneberg 2012) as “truth” and obtain a technical error of measurement using the assessors’s body heights and the formula given in Section 3.2.1. As can be seen from Table 12, assessor 3 produced the lowest score. This surprising result may be explained by the fact that assessor 3 gained familiarity with the subjects as a consequence of spending a significant amount of time setting up the experimental images. We also believe that the laboratory experiments have been beneficial for training. His landmarks are therefore better placed, in particular for subjects standing with their upper limb at an angle from their body (e.g. when pushing a trolley). The other two assessors performed equally to each other, although with larger errors than assessor 3.

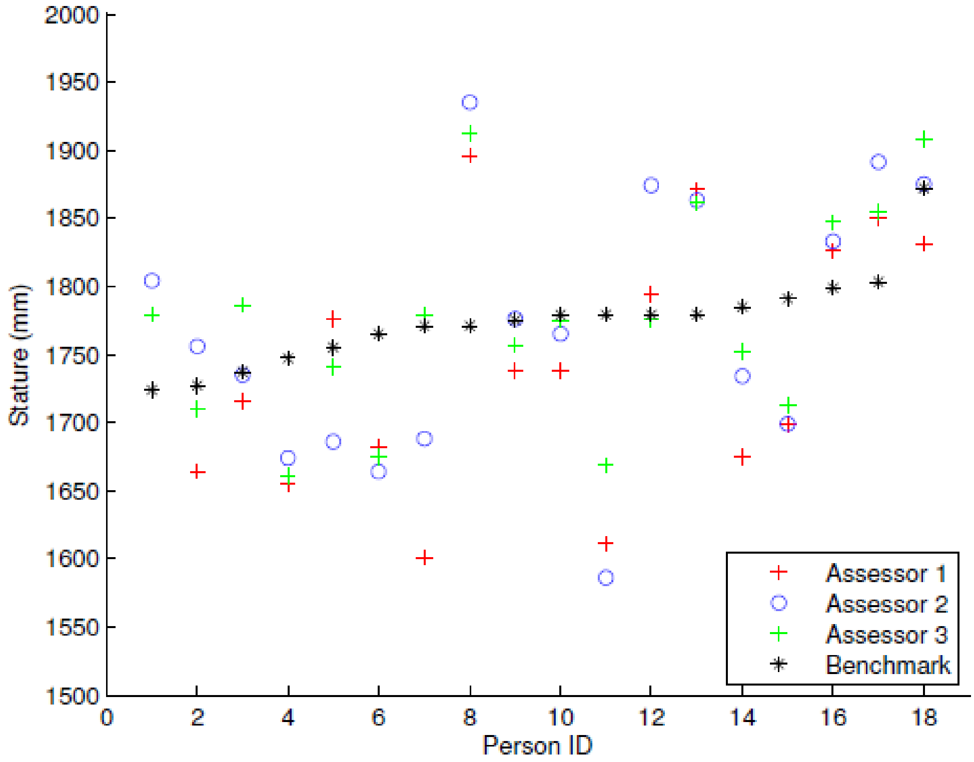
Table 12. Errors (mm) between predicted statures and (Scoleri and Henneberg 2012).

	Assessor 1	Assessor 2	Assessor 3
$\mathrm{TEM}_{\mathrm{truth}}$	60.8	61.2	45.3

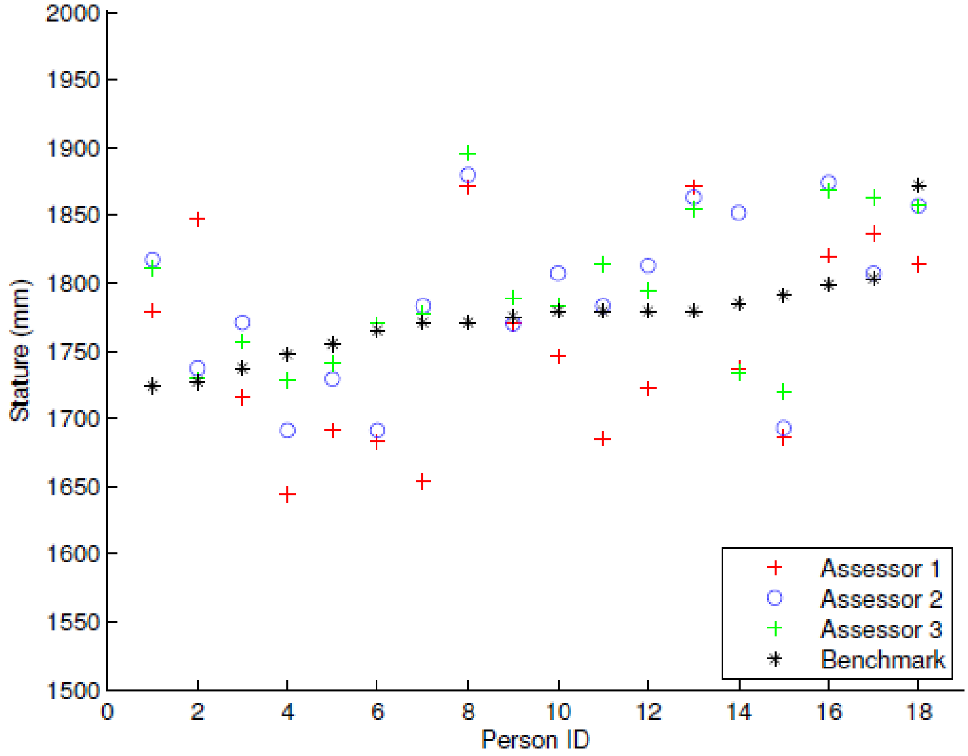
The assessors' inter-TEM and bias have been tested as per Sections 3.2.2 and 3.2.4, refer to Table 13. Given that assessors 2 and 3 have obtained better scores for $\mathrm{TEM}_{\mathrm{truth}}$, their inter-TEM and bias values are lower than those of other combinations of assessors. The negative bias values suggest that assessor 1 has obtained shorter statures on average compared to the other assessors.

Table 13. inter-TEM and bias (mm) for selections of assessors.

	Assessors			
	1 – 2	1 – 3	2 – 3	1 – 2 – 3
$\mathrm{TEM}_{\mathrm{truth}}$	34.9	40.3	32.3	36.0
Bias	-24.3	-30.8	-6.5	



(a) Test results



(b) Re-test results

Figure 9. Stature estimates shown on a macro scale. Benchmark stature originate from (Scoleri and Henneberg 2012).

Re-test results

A re-test session was organised to calculate the intra-observer TEM and observe the variations of other errors. So, the three assessors have repeated the point marking a second time for all subjects. Results are shown in Tables 14 and 15 with an accompanying graph in Figure 9(b).

Compared to TEMs in Table 12, the new values show improvement in landmark localisation and therefore stature estimation. Two of the assessors have scored TEMs equivalent to those in the laboratory experiments, which is encouraging. The average stature from (Scoleri and Henneberg 2012) is 1775 mm. Considering the worst error (assessor 1), this still gives an accuracy of about 97%. Looking at the intra-TEMs reveals that assessors have either become more precise in their marking or changed their approach (indicated by the large values). This is supported by the variations of predicted statures in the graphs of Figure 9. These intra-TEMs also suggest that assessors have a margin of progress.

Table 14. Quantification of errors (mm) after re-test.

	Assessor 1	Assessor 2	Assessor 3
$\mathrm{TEM}_{\mathrm{truth}}$	53.5	40.9	36.5
$\mathrm{TEM}_{\mathrm{intra}}$	41.0	48.3	34.1

Table 15. Inter-TEM and bias (mm) after re-test.

	Assessors			
	1 – 2	1 – 3	2 – 3	1 – 2 – 3
$\mathrm{TEM}_{\mathrm{inter}}$	46.4	47.1	28.1	41.5
Bias	-35.9	-39.8	-3.9	

According to Table 14, assessors 2 and 3 have close stature estimates (from $\mathrm{TEM}_{\mathrm{truth}}$), which implies that their landmark positions may be similar. In turn, this means they should produce a lower inter-TEM and bias compared to other pairs of assessors. This is indeed confirmed in Table 15. Results of assessor 1 are about as far apart from each of the other two assessors. The negative bias confirms that assessor 1 generally produces shorter statures as can be seen in Figure 9(b). Overall, the two graphs of Figure 9 show that assessors' predicted statures after re-test are less spread out, which is expected as they repeat the experiments.

Effects of garments

In the previous sections, $\text{TEM}_{\text{truth}}$ gives some global value with no distinction about the garment type. Using the statures obtained from the re-test experiments, we follow the same analytical process as in Section 3.2.5. The distribution of $\text{TEM}_{\text{truth}}$ per garment style and posture is summarised in Table 16. "Truth" is again taken as the predicted statures in (Scoleri and Henneberg 2012). The results for assessor 1 show that the error values are generally increasing with the complexity in garment style and subject posture. An inconsistency exists in row (d) where the error is abnormally large. This is rationalised by the fact that two of the three subjects in this category are pushing a trolley which creates an ambiguous situation, see Figure 6(a). The dactylion can be marked near the hand on the trolley or approximately half way down the thigh (according to the upper limb model of Figure 1). Assessor 1 decided on the former approach. The resulting upper limb lengths turn out much shorter than their actual lengths due to the bent elbow, hence the large TEM value. Assessors 2 and 3 opted for the latter approach. They have obtained greater lengths and consequently lower errors for this category and in row (e). Our current model would benefit from a multiple-part regression for resolving the ambiguity. Although this is not the intended focus of the present work, future extensions to accommodate the issue are possible and discussed in Section 4.

Table 16. $\mathrm{TEM}_{\mathrm{truth}}$ (mm) per assessor based on the garment style and subject posture. Categories (a) to (f) are described in the start of section 3.4.

	Assessor 1	Assessor 2	Assessor 3
(a)	27.3	13.1	8.5
(b)	37.0	39.5	43.5
(c)	44.6	52.3	41.2
(d)	74.3	23.4	16.5
(e)	63.7	33.5	6.5
(f)	58.8	62.5	63.2

Aside from these results, we also observe increasing error values (or very similar values) in other categories of assessors 2 and 3. Drawing special attention to row (e) of assessor 3, the error turned out very small. Investigation revealed that some of the images here are those that were used for developing the model. So, assessor 3 subconsciously gained familiarity with the subjects. Some level of progress is also expected from the first round of marking. This result is very powerful in that it indicates the extent to which the error may be decreased.

Table 17 presents the inter-TEMs for measurements which relate to specific garment types and subject postures. Most variations (largest errors) occur in the three hard-cases categories (rows d,e,f), especially in row (d) for the pairs 1-2 and 1-3. This is expected since assessor 1 has placed the landmarks most differently from the other two assessors for subjects in these classes. The inter-TEM for the pair 2-3 in row (d) is understandably smaller since these assessors have followed the same marking strategy. Performing Snedecor F-tests with N = 18 between rows (a)-(d), (a)-(e) and (a)-(f) of pairs 1-2 and 1-3 confirm that there are statistically significant differences at the 0.05 significance level.

Table 17. Inter-TEM (mm) for groups of assessors based on the garment style and subject posture.

	Assessors			
	1 – 2	1 – 3	2 – 3	1 – 2 – 3
(a)	30.5	24.4	12.5	23.7
(b)	24.9	26.6	23.5	25.1
(c)	31.9	29.5	12.7	26.1
(d)	68.9	80.7	19.8	62.3
(e)	58.4	66.7	33.3	54.7
(f)	47.2	12.2	48.7	39.8

Overall, although results have improved from the first experiments, substantial differences are present. Indeed, even when considering the best results (for assessors 2-3), squared inter-TEM ratios exceed cut-off F-test values. This proves that there still exists significant variations in the measurements and therefore all errors should be reducible further. This analysis agrees with the conclusion from examining the intra-TEMs (Table 14).

The bias between garments (which includes subject posture) is calculated by selecting statures from the first category (easy-shirt) and comparing to other categories for all assessors, see Table 18. Clearly, the effects on stature estimation become increasingly important as the garment style gets thicker and the subject posture is more complicated. The inconsistency for the pair (a)-(d) simply reflects the incorrect measurements of assessor 1 as seen in row (d), column 1 of Table 16.

Table 18. Bias (mm) between different garment style and subject posture.

	(a)-(b)	(a)-(c)	(a)-(d)	(a)-(e)	(a)-(f)
Bias	+13.2	+32.1	+70.3	+45.2	-40.7

Discussion

The experiments have shown that, despite landmark positioning errors, the accuracy of height estimates from real-life CCTV images can be commensurate with, if not surpass, the expected 95% accuracy of height reconstruction from direct measurement of skeletal remains. This has been a long challenge. The use of the upper limb length compared to the head height (Scoleri and Henneberg 2012) has decreased user interaction to a single action or two instead of extensive repeats to guarantee equivalent accuracy. This claim is supported by the results of assessors coming from a range of backgrounds in anatomy and computer vision. Variations between assessors suggest that some training in photoanthropometry is beneficial to reduce marking errors—*Practice makes perfect*.

We infer from the results that at present the weakest component in our procedure lies in the regression model, not the human factor. Large international databases of body measurements exist through the Civilian American and European Surface Anthropometry Resource (CAESAR) project (2002). We intend to use these anthropomeasures to construct improved regression equations. This would include multiple regressions from various body dimensions to refine our current upper limb model and combine it with other ones such as (Scoleri and Henneberg 2012) to allow for a more complete characterisation of human beings.

This research has permitted us to learn about the extent of errors involved in precise body part measurement from real-life surveillance imagery. The lessons may now be applied to enhance body-part acquisition from an automated detector or tracker. In addition, this knowledge could help in improving the recognition rate when matching subjects in uncontrolled scenes to ideal CAESAR data (2002).

Conclusion

This paper has examined the effects of various types of garments on human stature estimation from images. To this end, we have developed a procedure whereby the upper limb length is first measured from the image and then stature is inferred by linear regression from it. Three assessors have experimentally marked upper limb points of subjects in both laboratory and real-life surveillance videos. In both scenarios, thicker garments and those with stripes produce higher inaccuracies in stature prediction, which is to be expected. Seven error measures are used to study the variations between obtained statures. The most valuable outcome is that errors are within the expected variance of the stature regressor. Thus, body heights from imaged upper limbs can be inferred with confidence that is no worse than the accuracy of reconstructed body parts in routine skeletal forensic work.

References

Aggarwal, J, Cai, Q 1999, ‘Human motion analysis: A review’, *Computer Vision Image Understanding*, vol. 73, pp. 428–440.

Barron, C, Kakadiaris, I 2001, ‘Estimating anthropometry and pose from a single uncalibrated image’, *Computer vision and image understanding*, vol. 81, no.3, pp. 269–284.

BenAbdelkader, C, Cutler, R, Davis, L 2002, ‘Person identification using automatic height and stride estimation’, *Proceedings of International Conference on Pattern Recognition*, IEEE, pp. 377–380.

BenAbdelkader, C, Davis L 2006, ‘Estimation of anthropomeasures from a single calibrated camera’, *Proceedings of International Conference on Automatic Face and Gesture Recognition*, IEEE, South Hampton, pp. 499–504.

BenAbdelkader, C, Yacoob, Y, 2008, ‘Statistical body height estimation from a single image’, *Proceedings of International Conference on Automatic Face and Gesture Recognition*, IEEE, Amsterdam, pp. 1–7.

Bertillion, A 1886, *Identification anthropometriques. Methode nouvelle de determination de L’iidentite individuelle*, Masson, Paris.

Bertillon, A 1890, *La photographie judiciaire, avec un appendice sur la classification et l’identification anthropome´triques*, Gauthier-Villars, Paris.

Blau, S, Ubelaker, D 2009, *Handbook of Forensic Anthropology and Archaeology*, Left Coast Press, California.

Burton, A, Wilson, S, Cowan, M, Bruce, V 1999, ‘Face Recognition in Poor-Quality Video: Evidence from Security Surveillance’, *Psychological Science*, vol.10, pp. 243-248.

Brooks, M, Chojnacki, W, Gawley, D, Hengel, A 2001, ‘What value covariance information in estimating vision parameters’, *Proceedings of International Conference Computer Vision,* IEEE, Vancouver, pp. 302–308.

Byers, S, Myster, S 2005, *Forensic Anthropology*, Pearson, Boston.

Cao, D,Chen, C, Piccirilli, M, Adjeroh, D, Bourlai, T, Ross, A 2011, ‘Can facial metrology predict gender?’, *Proceedings of the International Joint Conference on Biometrics*, IEEE, Washington DC, pp. 1–8.

Civilian American and European Surface Anthropometry Resource (CAESAR) database, 2002, http://store.sae.org/caesar/.

Chen, S, Mau, S, Harandi, M, Sanderson, C, Bigdeli, A, Lovell, B 2010, ‘Face recognition from still images to video sequences; A local-feature-based framework’, *Journal on Image and Video Processing*, vol. 2011, pp. 1-14.

Collins, R, Gross, R, Shi, J 2002, ‘Silhouette-based human identification from body shape and gait’, *Proceedings of International Conference on Automatic Face and Gesture Recognition*, IEEE, Washington DC, pp. 366–371.

Criminisi, A 1999, ‘Accurate visual metrology from single and multiple uncalibrated images’, Ph.D. thesis, University of Oxford, Dept. Engineering Science, D Phil . thesis URL http://www.robots.ox.ac.uk/_vgg.

De Angelis, D, Sala, R, Cantatore, A, Poppa, P, Dufour, M, Grandi, M, Cattaneo, C 2007, ‘New method for height estimation of subjects represented in photograms taken from video surveillance systems’, *International Journal of Legal Medicine*, vol. 121, no. 6, pp 489–492.

Edelman, G, Alberink, I, Hoogeboom, B 2010, ‘Comparison of the performance of two methods for height estimation’, *Journal of Forensic Science*, vol. 55, no.2, pp. 358-365.

Fan, J, Yu, W, Hunter L 2004, *Clothing Appearance and Fit: Science and Technology*, Woodhead Publishing, Cambridge.

Frith, H, Gleeson, K 2004, ‘Clothing and embodiment; men managing body image and appearance’, *Psychology of men and masculinity*. vol.4, pp. 40–48.

Henneberg, M 2007, ‘Facial mapping, body mapping and the duties of an expert witness’, http://www.lawlink.nsw.gov.au/lawlink/pdo/ll_pdo.nsf/pages/PDO_facialapping.

Henneberg, M 2008, ‘Expert witness in a courtroom: The Australian experience’ In: Oxenham M (ed), *Forensic approaches to death disaster and abuse*. Australian Academic Press, Queensland, pp. 307-319.

Johnson, M 1991, ‘Appearance modifying garments’, U.S. Patent. Patent no.: 5,065, 458.

Kovesi, P 2009, ‘Video surveillance: Legally blind?’, *Proceeding of International Conference on Digital Image Computer Technology and Applications*, IEEE, Melbourne, pp. 204–211.

Krishan, K, Vij, K 2007, ‘Diurnal variation of stature in three adults and one child’, *Anthropologist*, vol. 9 no. 2, pp. 113–117.

Lee, S, Choi, J 2010, ‘A single-view based framework for robust estimation of height and position of moving people’, *Proceedings of the Pacific Rim Conference on Advances in Image and Video Technology*, IEEE, Las Vegas, pp. 562–574.

Lucas, T, Kumaratilake, J, Henneberg, M 2014, ‘The extent to which garments affect the body shapes of males from faceless CCTV images’, *Journal of Biological and Clinical Anthropology,* vol. 71. pp. 259-274.

McCoppin, R, Rizki, M, Tamburino, L, Freeman, A, Mendoza-Schrock, O 2012, ‘The effects of clothing on gender classification using LIDAR data’, *National Aerospace and Electronics Conference (NAECON)*, IEEE, Dayton, pp. 134–139.

Martin, R, Saller, K 1957, *Lehrbuch der Anthropologie*, Gustav Fischer Verlag, Stuttgart.

Nguyen, N, Hartley, R 2012, ‘Height measurement for humans in motion using a camera: A comparison of different methods’, *Proceedings of International Conference on Digital Image Computer Technology and Applications*, IEEE, Fremantle, pp. 1–8.

PETS: Performance Evaluation of Tracking and Surveillance benchmark data, 2007, http://pets2007.net/ (2007).

Rudd, N, Chattaraman, V 2006, ‘Preferences for aesthetic attributes in clothing as a function of body image, body cathexis and body size’, *Clothing and Textiles Research Journal*, vol. 24, no.1, pp. 46–61.

Scoleri, T 2010, ‘Video metrology without the image-to-ground homography’, *Proceedings International Conference on Digital Image Computer Technology and Applications*, IEEE, Sydney , pp. 335–342.

Scoleri, T, Henneberg, M 2012, ‘View-independent prediction of body dimensions in crowded environments’, In *International Conference on Digital Image Computing Techniques and Applications,* DICTA, Fremantle, pp. 1-8.

Thompson, P, Mikellidou, K 2011, ‘Applying the Helmholtz illusion to fashion: horizontal stripes won’t make you look fatter’, *Perception,* vol. 2, no.1, pp. 69–76.

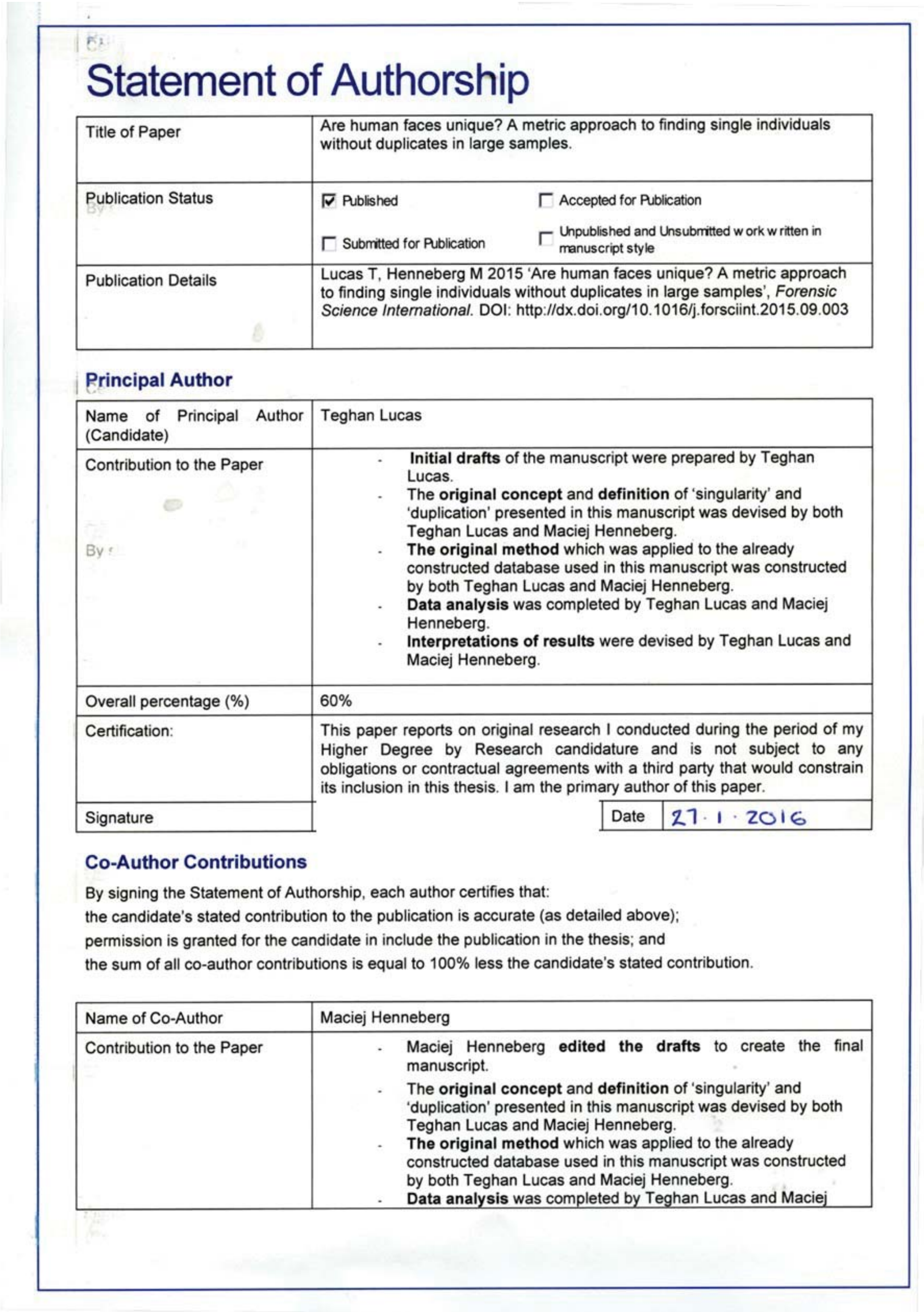
Ulijaszek, S, Lourie, L 1994, Intra- and inter-observer error on anthropometric measurement, Ulijaszek, S. J. and Mascie-Taylor, C.G.N. (ed), *Anthropometry: The individual and the population*, Cambridge University Press, Cambridge.

Voss, L, Bailey 1997, ‘Diurnal variation in stature: is stretching the answer?’, *Archives of Disease in Childhood*, vol.77, pp.319–322.

Willey, P 2009, Stature Estimation, in Blau and Ubelaker (ed.), *Handbook of Forensic Anthropology and Archaeology*, Left Coast Press, Walnut Creek.

Yoshino, M, Noguchi, K, Atsuchi, M, Kubota, S, Imaizumi, K, David, C, Thomas, L, Clement, J 2002, ‘Individual identification of disguised faces by morphometrical matching’, *Forensic Science International*, vol.127, pp. 97–103

Statement of Authorship Manuscript 2:



Statement of Authorship

Title of Paper	Are human faces unique? A metric approach to finding single individuals without duplicates in large samples.
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Lucas T, Henneberg M 2015 'Are human faces unique? A metric approach to finding single individuals without duplicates in large samples', *Forensic Science International*. DOI: http://dx.doi.org/10.1016/j.forsciint.2015.09.003

Principal Author

Name of Principal Author (Candidate)	Teghan Lucas
Contribution to the Paper	- **Initial drafts** of the manuscript were prepared by Teghan Lucas. - The **original concept** and **definition** of 'singularity' and 'duplication' presented in this manuscript was devised by both Teghan Lucas and Maciej Henneberg. - **The original method** which was applied to the already constructed database used in this manuscript was constructed by both Teghan Lucas and Maciej Henneberg. - **Data analysis** was completed by Teghan Lucas and Maciej Henneberg. - **Interpretations of results** were devised by Teghan Lucas and Maciej Henneberg.
Overall percentage (%)	60%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.

Signature		Date	27 · 1 · 2016

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

the candidate's stated contribution to the publication is accurate (as detailed above);

permission is granted for the candidate in include the publication in the thesis; and

the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Maciej Henneberg
Contribution to the Paper	- Maciej Henneberg **edited the drafts** to create the final manuscript. - The **original concept** and **definition** of 'singularity' and 'duplication' presented in this manuscript was devised by both Teghan Lucas and Maciej Henneberg. - **The original method** which was applied to the already constructed database used in this manuscript was constructed by both Teghan Lucas and Maciej Henneberg. - **Data analysis** was completed by Teghan Lucas and Maciej

	Henneberg. - **Interpretations of results** were devised by Teghan Lucas and Maciej Henneberg.		
Signature		Date	27.1.16

Please cut and paste additional co-author panels here as required.

2. Are human faces unique? A metric approach to finding single individuals without duplicates in large samples.

Teghan Lucas[a*] and Maciej Henneberg[a]

[a] Biological and Comparative Anatomy Research Unit, Medical School, University of Adelaide, Adelaide 5005, Australia.

Submitted to Forensic Science International on the 22nd of August 2014, accepted on 8th September. **Published** on 30th September.

[*]Corresponding author.

Teghan Lucas

Medical School North Building, Frome Road, The University of Adelaide, South

Australia 5005, Phone: +61 8 8313 6324. Email address: teghan.lucas@adelaide.edu.au

Context

Personal identification in the forensic sciences rely on the notion that each individual is ‘unique’. Without individuality, prosecutions of guilty people based on biological evidence would cease to exist. Many studies have shown that different biological characters have a low probability of being repeated in the human population, these include: DNA (Jeffreys et al. 1985), fingerprints (Jain et al. 1997; Jain, Prabhakar and Pankanti 2002), bite marks (Mahaja et al. 2012) elbow prints (Oatess 2000) and ear prints (Meijerman et al. 2004). The increased use of closed circuit television video (CCTV) systems is reintroducing anthropometric measurements as a forensic biological marker (source), although their usefulness is often debated. In order to prosecute a person based on biological evidence, it must be shown beyond reasonable doubt that a biological characteristic (trace) left at the crime scene is a ‘match’ with the suspect (source). Although the methodology supporting these studies is sound, many researchers debate the term ‘unique’, claiming that it cannot be proven (Cole, 2009).

The aims of this study were to introduce the term ‘singularity’ which is defined as ‘a situation when only one individual in a specific population has a particular set of characteristics’. To achieve this aim, a large anthropometric survey of facial measurements was used to search for the presence of duplicates (two or more individuals who match on a specified set of characteristics). A secondary aim was to calculate probabilities of finding two or more duplicate individuals in the world.

Studies that use the human face for identification are doing so assuming that no two individuals have the same set of characteristics. After an extensive literature search and analyses, it was found that no references were provided for studies that claim the face is ‘unique’ or sufficient to use to isolate an individual from a large sample. Thus, this research aimed to provide such a reference while introducing the term ‘singularity’, which would avoid any ambiguity associated with the word ‘unique’. This research showed that singularity can be achieved in a large population of anthropometric measurements. It also showed that probabilities of finding a duplicate individual are so low that no two human beings in the world have the same face/head measurements.

Singularity is a new concept and there was some difficulty in illustrating its meaning to reviewers of the paper. However, this was overcome by improving the definitions.

The previous manuscript ‘Effects of garments on photoanthropometry of body parts: Application to stature estimation’, demonstrated that measurements can reliably be taken from images. This manuscript aims to provide theoretical information about the human face (which may be measured from images by other studies) and its usefulness in isolating an individual in a population.

Abstract

In the forensic sciences it is inferred that human individuals are unique and thus can be reliably identified. The concept of individual uniqueness is claimed to be unprovable because another individual of same characteristics may exist if population size were infinite. It is proposed to replace “unique” with “singular” defined as a situation when only one individual in a specific population has a particular set of characteristics. The likelihood that in a population there will be no duplicate individual with exactly the same set of characteristics can be calculated from datasets of relevant characteristics. To explore singularity, the ANSUR database which contains anthropometric measurements of 3982 individuals was used. Eight facial metric traits were used to search for duplicates. With the addition of each trait, the chances of finding a duplicate were reduced until singularity was achieved. Singularity was consistently achieved at a combination of the maximum of seven traits. The larger the traits in dimension, the faster singularity was achieved. By exploring how singularity is achieved in subsamples of 200, 500 etc. it has been determined that about one trait needs to be added when the size of the target population increases by 1000 individuals. With the combination of four facial dimensions, it is possible to achieve a probability of finding a duplicate of the order of 10^{-7},while, the combination of 8 traits reduces probability to the order of 10^{-14}, that is less than one in a trillion.

Keywords: Forensics; Human variation; Identification; Singularity

Introduction

In the forensic sciences the process of identification relies upon a finding that a trace left at the crime scene (or other location relevant to the investigation) and the suspected source, often an object or a person, correspond to each other in essential characteristics. In brief, they match. In order to make a match between a trace and a source that allows conclusion of identification, the probability of finding a duplicate trace or a duplicate source must be negligible. If this probability equals zero the match is unique (Page, Taylor and Blenkin 2011). However, it can be argued that unless the whole world is included in searches for duplicates, the probability cannot be firmly established as zero (Cole, 2009). This makes the term "unique" debatable. In real situations populations of traces and of sources are not infinite, thus probabilities of finding a duplicate can only approach zero. In those cases, when the probability of finding a duplicate is less than what the population size predicts, the actual match between the trace and the source is a single occurrence (Page, Taylor and Blenkin 2011). We propose to call this match a "singularity". In everyday terms it could be called a "unique correspondence of trace and source essential characteristics", but the ambiguity of the term "uniqueness" remains a problem. We define singularity as the correspondence between essential characteristics of the trace and the source that in a given population has a probability of occurrence less than that predicted from random combination of characteristics of the trace and the source. This probability predicts that no duplicate of the trace or of the source can be found in the given population.

The term "singular" as defined here is free from the ambiguity of the word "unique" and thus may be more appropriate to use in forensic statements and court proceedings. Unlike statements saying that a particular individual is unique, the statement that the individual is singular in a defined population is easily testable both empirically and in court proceedings.

The lower the probability of finding a duplicate trace the more reliable the evidence is considered. Some widely recognised claims of unique traces left by humans are: DNA (Jeffreys et al. 1985) fingerprints (Jain et al. 1997; Jain, Prabhakar and Pankanti 2002) bite marks (Mahaja et al. 2012). The lesser known traces of this kind include: elbow prints (Oatess 2000) ear prints (Meijerman et al. 2004) lip prints (Mishra, Ranganathan and Saraswathi 2009) and behavioural characteristics such as gait (Bouchrika et al. 2011) and handwriting (Crane 1999). Uniqueness of these traces is claimed on the principle that only one individual would be a source of such trace. This is to be criticised because, theoretically, another individual could be born with the same essential characteristics if we wait long enough. In practice, what is considered uniqueness of such traces is actually a singularity since populations from which those traces emanate are limited by time and space and thus are of a finite size.

The word ‘traces’ can refer to light rays producing changes on photographic film or on light sensitive digital photograph chips, that is to images. In recent years photographic traces have become increasingly popular as forensic evidence. In cases of morphological analyses, an expert witness specialising in the field of biological anthropology is often called upon to provide evidence of a match or mismatch between an image of an individual (a trace) and a suspect (a source) based on anatomical similarities of morphological traits. The current method for analysing image based traces is to use categorical scales of morphological traits (Henneberg 2007: 2008, Rosing 2006). For example, body height may be described as short, medium or tall. This method has been highly criticised on the grounds that it is not accurate and reliable (Edmond 2008, Edmond et al. 2009, Edmond 2010). Attempts have been made to address criticisms by attempting to take measurements from images, however, this research is still in progress (Scoleri and Henneberg 2012, Scoleri, Lucas and Henneberg 2014).

Over the last few decades, facial recognition systems have been increasingly researched due to the proliferating use of images for identification. Some studies explain that the face is used as humans are good at recognising facial features (Shi, Samal and Marx 2006, Burton et al. 1999), others say that obtaining images of the face is cheap and non-invasive (Jafri and Arabnia, 2009). According to Jafri and Arabnia (2009) facial recognition is used for two primary tasks, verification (one to one matching) and identification (one to many matching).

There are a number of different scenarios where face recognition can be used including: passports, drivers licences, security, surveillance etc. Some factors that make identification from images difficult include: illumination, facial expression, pose, distortions and pixelation (Chen et al. 2010). Therefore research has concentrated on eliminating the confounding effects from images when making an identification.

There are two ways to analyse facial traits, descriptives and metrics. The use of descriptive traits involves adjectives such as ‘wide’ and ‘curved’ to categorise facial features such as the nose. Metric traits most commonly involve measuring the distances between specific points on the face. Theoretically, any descriptive trait, such as, for instance, face shape, can be converted to a metric one by taking measurements of its constituent properties such like the width, the height, the curvature, angles between its parts etc. As mentioned earlier, descriptives are not considered as a reliable method of evaluation, especially in court proceedings (Edmond 2008, Edmond et al. 2009, Edmond 2010). Therefore, many facial recognition systems and methods have concentrated on moving towards metrics (Cattaneo et al. 2012) or a combination of metrics and descriptives (Klare and Jain 2010; Ritz-Timme et al. 2011). Facial recognition methods which use metrics alone or metrics and descriptives report high accuracy rates (>95%) (Turk and Pentland 1991; Belhumeur, Hespanha and Kriegman 1997). However, each method of finding a match between a photograph and a person is tested on a small sample of subjects ie. less than one hundred. Many photographs are used and a large number of measurements taken, but the number of people appearing on these photographs is limited. Small sample size automatically assures that no duplicate matches are found within the sample. These studies are based on the assumption that a face is unique, but very few justify this assumption quantitatively. In those studies that mention ‘uniqueness’ of the face, the term is not referenced nor defined. Therefore it seems that many people believe that the face is adequate to be used as a biometric tool, without it being sufficiently studied, especially when using metrics. A number of papers have identified the lack of knowledge in this area (Spaun 2007: 2009, Klare and Jain 2010).

The aim of this study is to investigate whether or not two or more faces within a specified population have the same combination of several measurements. A secondary aim is to calculate the probability of not finding more than one face with same measurements (a duplicate) using a defined number of measurements in order to find the minimum number of facial dimensions needed to achieve singularity. We are not aiming here to investigate the accuracy with which measurements can be taken of various images and how such accuracy may influence findings of singularity. We only aim to introduce the principle of finding singularity that can be applied to any traces or sources, while the precision of its application requires a separate discussion of measuring techniques that may differ from case to case.

Materials and Methods

The U.S Army Anthropometric Survey (ANSUR) database is a result of an anthropometric survey conducted in 1988 of U.S military personnel. The dataset contains 132 manually measured dimensions of the human body and head. The sample consisted of 1774 men and 2208 women aged 17 – 51 years. This dataset was chosen as it is a sample of anthropometric measurements covering a range of variation of facial features. Even though the survey was conducted in 1988, it is still valid for the purposes of this study because the human face has not varied statistically within that time. Details of this study are described in ANSUR (1998). Initially males and females were analysed separately to see if sex influenced the numbers of metric traits needed to have no duplicates. Males and females were then combined to increase sample size and because sexual dimorphism accounts for only 25% variation in major measurable characters (Henneberg 2010). No further separation of the dataset (i.e. population of origin) was included as upwards of 95% of variation occurs between two randomly selected individuals rather than between individuals of different populations (Cavalli-Sforza and Bodmer 1971)

A team of 22 individuals conducted all measurements on the sample. The team was trained in anthropometry over a four week period. During this time, each member of the team was allocated specific measurements to learn and repeat continuously. These dimensions were then measured on the 3982 participants during data collection. By allocating specific measurements to each measurer, the measuring team aimed to reduce measurement errors. Measurement errors were calculated and reported alongside the database. Measurement errors ranged between 2.2% and 2.4% which is very small. For the purposes of this paper, measurement errors will not affect the results.

The following face/head measurements were used (Figure 1): *Bitragion submandibular arc* – The surface distance between the right and left tragion across the submandibular landmark at the juncture of the jaw and the neck was measured with a measuring tape . *Bizygomatic breadth* – The maximum horizontal breadth of the face between the zygomatic arches was measured with a spreading caliper. *Ear length* – The length of the right ear is measured with a sliding caliper from its highest to lowest points on a line parallel to the long axis of the ear. *Ear protrusion* – The horizontal distance between the mastoid process and the outside edge of the right ear at its most lateral point (ear point) is measured with a sliding caliper. *Head breadth* – The maximum horizontal breadth of the head above the ears is measured with a spreading caliper. *Head circumference* – a measuring tape is used to measure the maximum circumference of the head above the supraorbital ridges and ears. *Head Length* – the distance from the glabella landmark between the brow ridges to opisthocranion is measured with a spreading caliper. *Interpupillary distance* – a pupillometer is used to measure the distance between the centres of the right and left pupils. The dimensions are illustrated in Figure 1. The dimensions that are not influenced by facial expression were chosen.

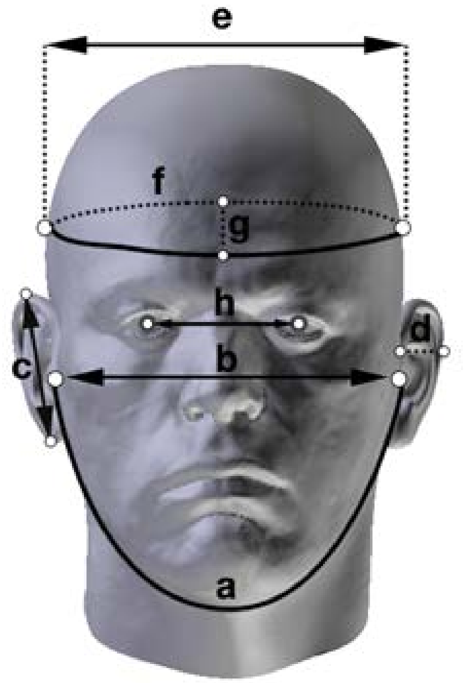


Figure 1: Anterior view of the head/face showing: (a) bitragion submandibular arc (b) bizygomatic breadth (c) Ear length (d) Ear protrusion (e) Head breadth (f) head circumference (g) head length (h) Interpupillary distance.

Duplication is defined here as '*a situation where another trace is found matching a given trace'* since we have at our disposal only measurements (traces) of participants in the survey. In real forensic situations dimensions of a trace can be compared with dimensions of the actual person – the source. Depending on the number of traits measured on a trace and the accuracy with which they are measured, the ease of finding a duplicate will vary. If only a combination of two metric traits is used, it will be easier to find duplicates than when more dimensions will be used to characterise traces. As long as dimensions are not perfectly correlated with others ($r<1.00$), adding a dimension should reduce the number of possible duplicates. In a defined sample, when a larger number of combinations of metric traits (traces) is used, a situation can be reached where no duplicates will be found. In the case of identification, a trace left behind by a particular individual can be analysed and described by a number of metric traits. This trace can be compared to the traces produced by the same measuring methods of any number of possible suspects from a defined population. . If only one duplicate of the source is found in this sample, an ‘identification’ can be made.

All anthropometric measurements in the ANSUR database are reported to the nearest millimetre. IBM SPSS statistics 20 was used to search for duplicate cases within the sample. To put it simply, duplicates occur when exact values of all metric traits (traces) selected for analysis are found in more than one person. Sorting of cases was done stepwise, adding one dimension at a time to the previous dimension(s) and noting how many duplicates are found with the combination of that number of traits. This procedure was continued, increasing the number of traits until no duplicate cases were identified within the sample. For example, the first measurement analysed was bitragion submandibular arc, the number of duplicate cases identified in a sample of males was 1686. Then bizygomatic breadth was added as a second trait reducing the number of duplicate cases to 908. The third trait added was ear length, and then the number of duplicates fell down to 122. This was continued by adding ear protrusion, and then head breadth at which point no duplicates were found.

Polynomial regressions were used to study the shape of relationships between numbers of traits considered and numbers of duplicates. Those regressions allowed us to extrapolate results beyond sample sizes available.

Results

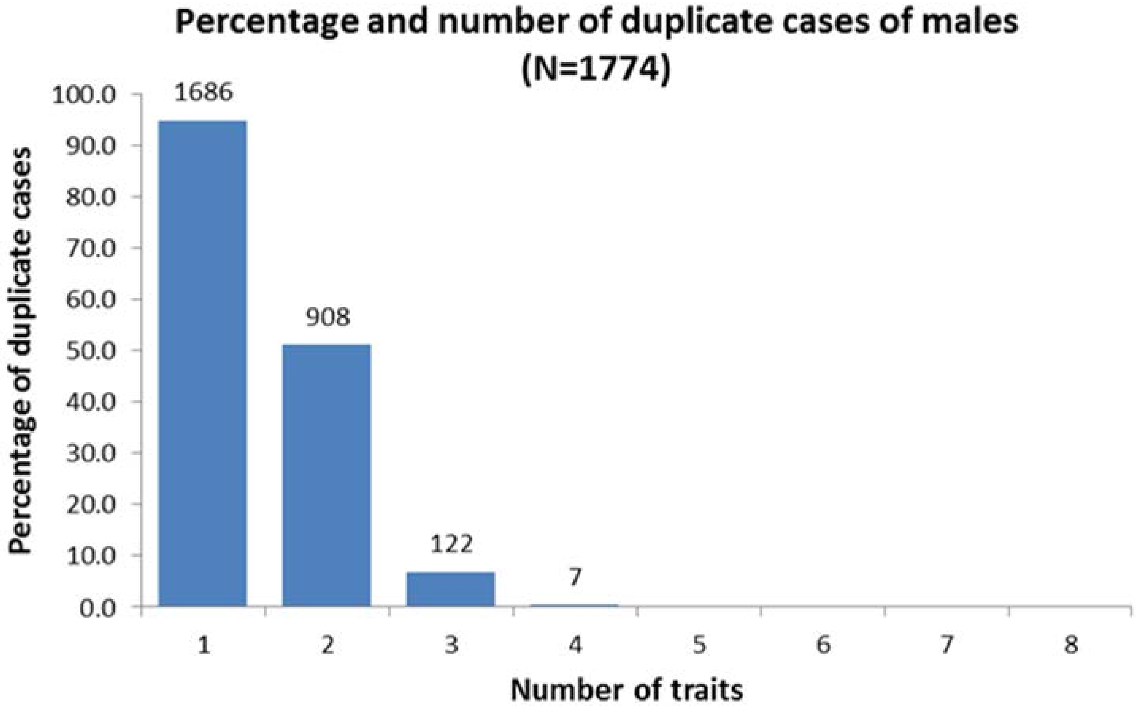
Table 1 shows the metric traits used with their means for males and females combined, the order of the list changes throughout analysis (as discussed) to establish any effect the order of traits may have on the outcome.

Table 1: List of metric traits and their means (mm) for the entire sample of 3982 females and males.

Metric traits	**Mean**
Bitragion submandibular arc	288.6
Bizygomatic breadth	135.4
Ear length	62.0
Ear protrusion	22.9
Head Breadth	147.7
Head circumference	555.8
Head length	191.6
Interpupillary distance	63.4

Note: this list is in alphabetical order as presented in the ANSUR database. The order of metric traits has been changed throughout the analyses as indicated.

The sample was divided by sex to determine how many traits are needed to have no duplicate cases in each sex. The order of traits chosen was alphabetical as found in the original ANSUR files. Figure 2 shows the percentage and number of duplicate cases for males (N=1774) and females (N= 2208). There is a rapid decline in the number (and thus percentage) of duplicate cases with the addition of each trait to the previous trait(s). In order to have no duplicate cases in males only 5 traits are needed, whereas in females, 6 traits are needed. Thus there is little difference between the sexes in the number of traits required to reduce number of duplicate cases to zero.



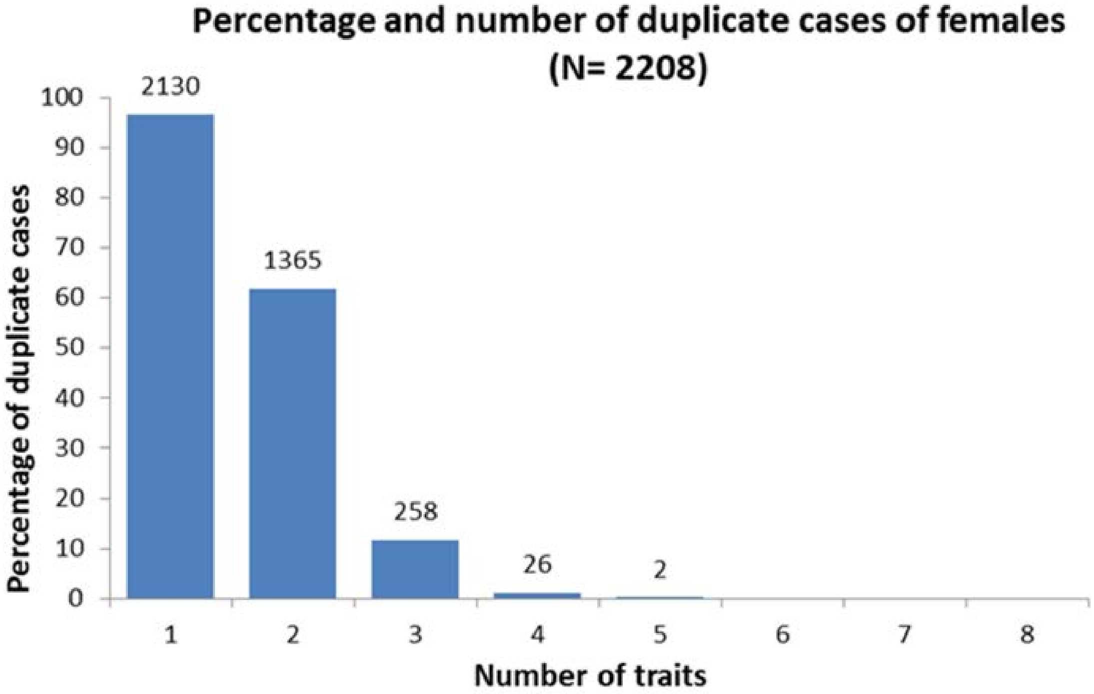


Figure 2: The percentage and number of duplicate cases for males and females. The order of metric traits chosen is alphabetical.

Due to there being little difference in the number of traits needed to find no duplicates and thus achieve a singularity in males and females separately, the sexes were combined for all further analyses and the order of individuals was randomised.

The order of metric traits has been changed from alphabetical to largest - smallest based on the mean of each trait measured (Table 1). The order was changed to establish any differences in the number of traits needed to achieve singularity when larger metric traits were chosen first. There is a large decrease in percentage of duplicate cases (59.4%) when the second trait (bitragion submandibular arc) is added to the first trait (head circumference). In contrast, when using alphabetical order, the largest decrease between the first (Bitragion submandibular arc) and second (bizygomatic breadth) traits was found in males at only 43.8% (Figure 2). It is important to note that the overall outcome of needing 5-6 traits to achieve singularity remains the same when the order of metric traits is changed. However, the rate of decrease in duplicate cases when using metric traits in largest-to-smallest order is faster, especially with the combination of two traits.

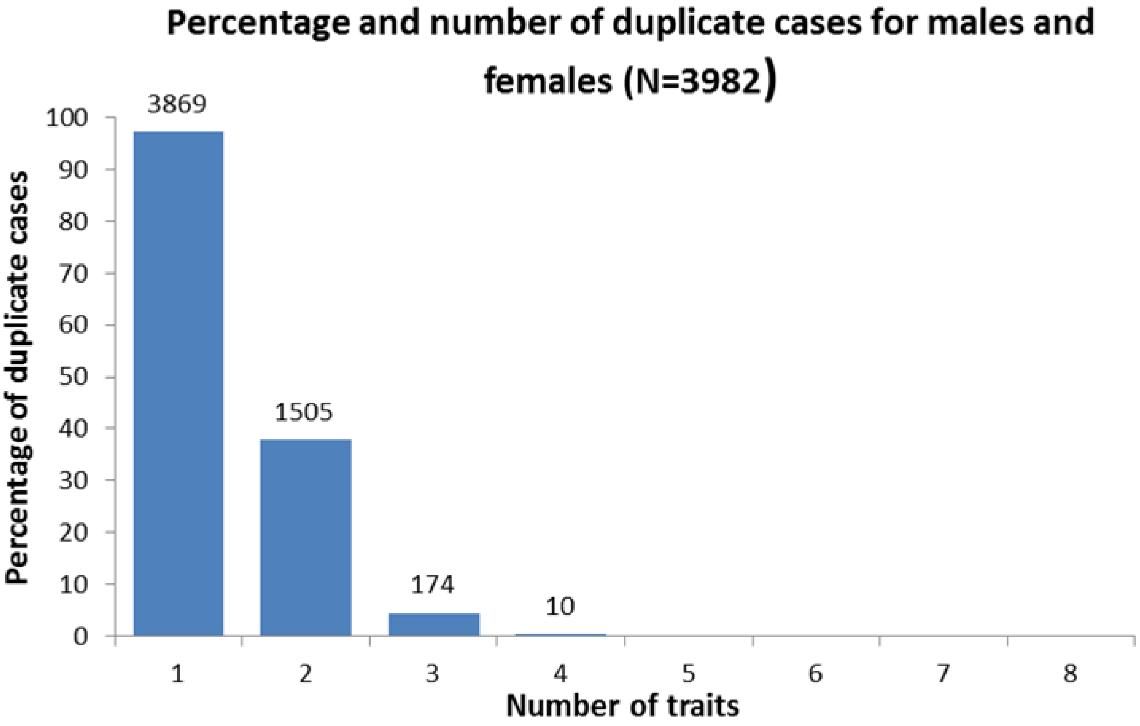


Figure 3. The percentage and number of duplicate cases for males and females. The order of metric traits chosen from largest to smallest.

The numbers of traits needed to achieve singularity was determined for sample sizes beginning at 200 subjects then adding 300, followed by adding 500 repeatedly at 1000, 1500 etc. until all subjects were included. This was conducted on the list of traits ranging from largest to smallest and then from smallest to largest as shown in Figure 4. The average and maximum number of traits needed to achieve singularity when the list is ordered largest to smallest is 5. The average number of traits needed to achieve singularity when the list is ordered smallest to largest is 6 with a maximum of 7. A smaller number of traits is needed to achieve singularity when the traits are larger in dimension since they have a greater range of numerical values. However, the average number of traits (5-6) needed to achieve singularity still remains consistent, no matter in what order the traits are included into calculation. The polynomial regression equation (Fig.4) indicates that approximately one more trait is needed for every 1000 people added to the sample in order to reduce the number of duplicate cases to zero. Polynomial regression was used to predict the number of traits needed in a sample of specific size exceeding that of the ANSUR database (Table 2). The number of possible combinations for each set of metric traits was calculated using a stepwise approach. The traits labeled 1-8 are seen here combined into sets in the order largest to smallest based on their means (Table 1). The number of units (range) is the number of possible metric values (mm) for a given metric trait, i.e. the difference between the minimum and the maximum size of the trait in full millimeters.

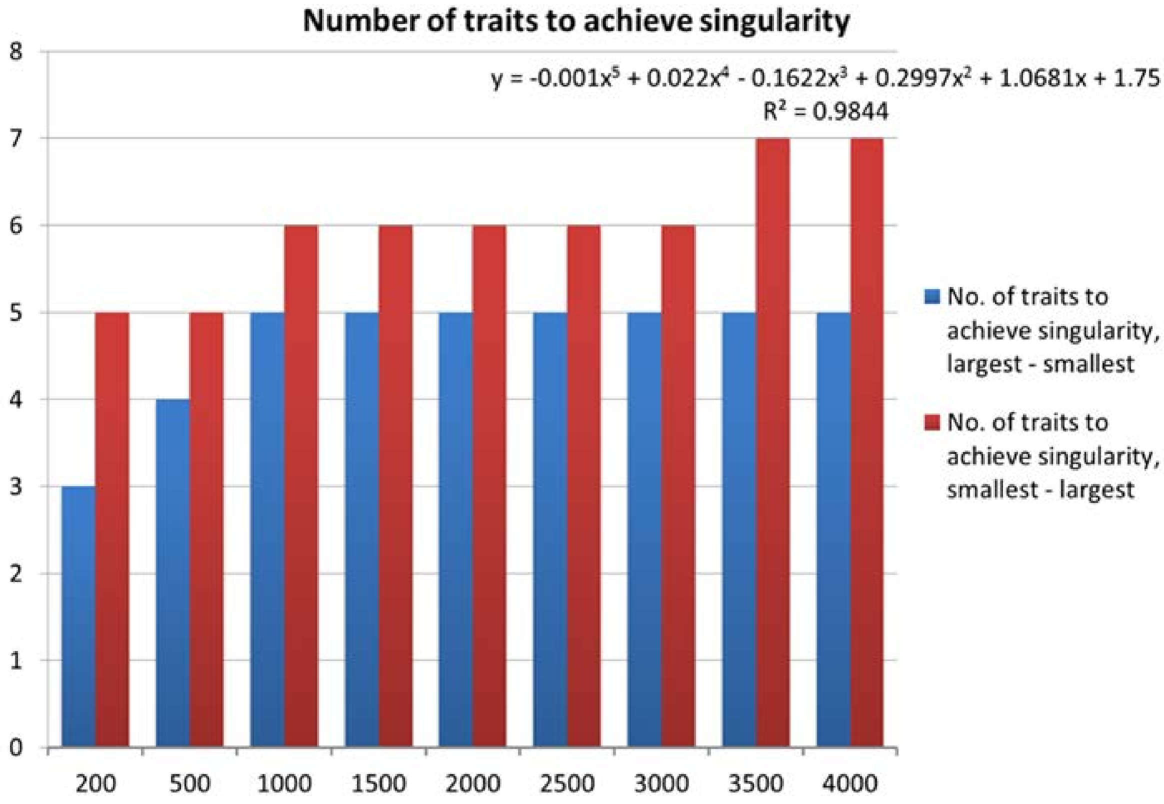


Figure 4: The number of traits needed to have no duplicates, the list of traits is ordered largest to smallest and smallest to largest.

Probabilities of sets of traits to occur together are difficult to calculate because of the intercorrelation between them. If there is a correlation between two traits then they cannot be viewed independently of one another and their joint probability of occurrence cannot be calculated by simple multiplication of individual probabilities.

This is the case with dimensions of the human face. Adjusted R squared values (or correlation coefficients) were calculated for each set in a stepwise approach by SPSS version 20. The R squared value was subtracted from one ($1-R^2$) to determine parts of traits that are not correlated and thus can be combined randomly. The total number of units possible in each set is a product of numbers of units in traits of the set reduced by their intercorrelation. The number of combinations is calculated by multiplying the $1-R^2$ value by the number of units in the first trait, then number of units in the second trait etc. The addition of each new trait to a set increases the number of combinations as long as the trait is not completely correlated with others, and thus decreases the possibility of finding another individual with the same set of traits.

Table 2: The calculation of probability of finding another duplicate within the sample, using eight metric traits, their respective ranges and R^2 values.

Traits	Number of units (range) in mm	R^2 correlation with preceding units	1-R^2	Multiplication of units	Number of combinations (1/p)	p
1	127			127	127	0.01
1,2	138	0.415	0.585	17526	10253	0.0001
1,2,3	62	0.780	0.220	1086612	239055	0.000004
1,2,3,4	47	0.638	0.362	51070764	18487617	0.00000005
1,2,3,4,5	44	0.736	0.264	2247113616	593237995	0.000000002
1,2,3,4,5,6	26	0.380	0.620	58424954016	36223471490	0.00000000003
1,2,3,4,5,6,7	35	0.336	0.664	2044873390560	1357795931332	0.0000000000007
1,2,3,4,5,6,7,8	27	0.229	0.771	55211581545120	42568129371288	0.00000000000002

Discussion

With the increased use of video surveillance systems for identification, using the face as an example to illustrate the method proposed in this paper seemed only fitting. Many people identify individuals by the face as it is believed to be the most variable part of the human body. The findings presented in this paper have a direct application to facial analysis by providing the probabilities of finding two individuals with the exact same facial metric characteristics. As a general rule the combination of traits is the key to reducing the number of duplicates within a specified population. Depending on the exact variation of each trait, and its intercorrelation with other traits, the number of traits needed to find no duplicates and thus achieve singularity varies. This number also increases with the target population size. The larger the range of each trait, the more possible combinations per set of traits and thus the faster a result of singularity is achieved. More variable, i.e. larger traits should always be chosen first, especially in sample sizes lower than 500 individuals. In cases of applying this theory to identification from images, the traits chosen depend on the quality of the images, the angle of the cameras, whether or not the person is covering their face and many more factors. Each case is different and must be treated accordingly. However, the general guidelines still apply and the theory remains the same. If , for whatever reasons, singularity is not achieved with 8 traits, simply adding more traits will increase the probability of achieving singularity by decreasing the probabilities of finding duplicates.

When comparing metric traits with traditional descriptive traits (which use categorical scales) metric traits have a significantly larger range of values since they include more intervals (eg. millimeters) than categorical scales. This fine gradation decreases the chances of finding a duplicate. Goldstein, Harmon and Lesk (1970) used descriptive traits with a maximum range of 5 categories per trait. Goldstein et al.’s study was able to isolate an individual from a set of 256 photographs. Not to complicate their analysis too much the authors assumed no correlation between traits, however they realised that at least some of them are correlated. Despite the simplicity of their method, Goldstein, Harmon and Lesk (1970) showed that in order to isolate an individual from a predicted population size of 10^7 it is sufficient to use 16 traits each with 5 categories. Although forensic sciences shift towards the use of traditional metric analysis for comparisons of images, the use of descriptives, however debated, is not without merit. Descriptive traits can be quantified to increase reliability. In Goldstein et al.'s case, only a limited number of categories per descriptive trait was used, thus the limited range produced inferior results. However, descriptive traits can have a significantly large range. For example, a mole above an individual's lip can be measured. Its height, width, diameter and location compared to anatomical landmarks can be calculated. Its colour can be measured metrically as a wavelength of light reflected from its surface. It may have an uneven shape that yields it to be measured in parts. Each case is different and each descriptive trait can be measured in endless ways that apply to a specific case. The same rules that apply to traditional metrics apply to descriptives, the larger the range the less chance of finding two individuals who match. However, in the case of descriptive traits, knowing the population frequencies of the occurrence of the traits would be useful in calculating probabilities.

In a study conducted by Kleinberg and colleagues (2007), it was found that proportions of 4 anthropometric measurements taken from photographs were not sufficient to make significant positive identification of 80 individuals. From this study it was deemed that 'anthropometry failed as an identification technique'. It should be discussed that in this study there were a limited number of measurements being taken and each of these measurements was on a very small area of the face. Two of the measurements included in the four were essentially the same measurement, the distance between the ectocanthions and the stomion on both the right and left side of the face. The variability of the measurements is very limited; therefore the ability to match an individual would have been limited. This study is discussed to reinforce the importance of variability and size of the traits chosen, the larger the size and variability the higher the probability of achieving singularity.

Here we have considered correlation of traits and used only the part of each trait which varies independently from other traits. By using traits measured in interval scales and considering intercorrelations among traits we are ensuring that the method is presenting the biological variation between individuals as accurately as possible. The accuracy with which traits are measured will also affect the results. When measurement errors are known, ranges of variation should be adjusted to take errors into account. For example, if the measurement error is 3 mm, then the distance of 33 mm will legitimately be a duplicate of any distance between 30 mm and 36 mm. Size of errors depends on the quality of images measured, ability to locate points between which measurements are taken and the accuracy of measuring instruments. These factors may differ from case to case and need to be discussed when presenting results of a particular case. A study conducted by Cummaudo and colleagues (2013) it was found that anthropometric landmarks and thus the distances between them are less reliable when taken from 2D photographs. This was also the case in a study conducted by Farkas (1994). Interobserver errors of measurements taken directly on the participants by well-trained measurers that we used in this paper were small and can be ignored for purposes of demonstration of the general principle of the proposed method. Were the method used in a particular case, errors of point location and measurements would have to be assed and taken into account when searching for possible duplicates.

In a court of law, the main question to be answered is, ‘what are the chances that two sources share the same trace’. In DNA analysis, the evidential weight of a match between crime stain profile (trace) and a suspect (source) is quantified by the ‘match probability’ (Jobling and Gill, 2004). In many cases the match probability is much smaller than the probability of presence of one individual in a population which the trace originated from, with the match probability going into the inverse of millions. Table 2 shows the metric trait version of a match probability represented by ‘p’, the probability of finding a match of the trace to more than one individual (source). In our sample, a combination of four metric traits already gives a probability lower than that found by Goldstein, Harmon and Lesk (1970), 10^{-8}, while a combination of eight traits lowers the probability to 10^{-14}, this is comparable with DNA (Jobling and Gill 2004). A fragment of the DNA molecule contains four different nucleotides that can occur in different combinations. The longer the fragment, the lower the probability that another fragment with the same pattern of nucleotides can be found. Like DNA the more information extracted from the trace, the lower the probability of finding a duplicate.

The current paper uses facial analysis as an example to support the theory of singularity, however, this theory in no way is limited to facial analysis. The concept of singularity and the method presented can be applied to all fields of forensic sciences which aim to match a trace to a source. Many biological examples have been presented in this paper, however. the concept can also be applied to matching a trace with a source left by an object, for example, the study of ballistics (Springer 1995), gunshot residues (Gibelli et al., 2010) tool marks (Cassidy 1980) tire prints (McDonald 1993) cut marks and tool marks (Reichs 1998) glass fragments (Rodriguez et al. 2008) and any other pattern matching of everyday objects (Jayaprakash 2013).

Conclusion

This paper introduces the concept of singularity to forensic identification. Within the specified population of the ANSUR database, no two individual faces matched one another on combinations of 5-8 metric traits. Probabilities of finding a duplicate of a face characterised by 8 traits exceeded inverse of the total population of the Earth making metric identification of faces as reliable as that of DNA. The same concept can be applied to identification based on measurements of human bodies or any traces of any objects.

References

Belhumeur, PN, Hespanha, JP, Kriegman, DJ 1997, ‘Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection’, *Pattern Analysis and Machine Intelligence,* vol.19, no.7, pp. 711-720.

Bouchrika, I, Goffredo, M, Carter, J, Nixon, M 2011, ‘On using gait in forensic biometrics’, *Journal of Forensic Sciences*, vol. 56, no.4, pp.882-889.

Burton, A, Wilson, S, Cowan, M, Bruce, V 1999, ‘Face Recognition in Poor-Quality Video: Evidence from Security Surveillance’, *Psychological Science,* vol.10, pp. 243-248.

Cassidy, FH 1980, ‘Examination of tool marks from sequentially manufactured tongue-and-groove pliers’, *Journal of Forensic Sciences*, vol. 25, no.4, pp. 796-809.

Cattaneo, C, Obertová, Z, Ratnayake, M, Marasciuolo, L, Tutkuviene, J, Poppa, P, Gibelli, D, Gabriel, P, Ritz-Timme, S 2012, ‘Can facial proportions taken from images be of use for ageing in case of suspected child pornography? A pilot study', *International Journal of Legal Medicine*, vol. 126, pp 139-144.

Cavalli-Sforza, LL, Bodmer, WF 1971, *The genetics of human population*. W.H. Freeman and Company, San Francisco.

Chen, S, Mau, S, Harandi, M, Sanderson, C, Bigdeli, A, Lovell, B 2010, 'Face recognition from still images to video sequences; A local-feature-based framework', *Journal on Image and Video Processing*, vol. 2011, pp. 1-14.

Cole, S 2009, 'Forensics without uniqueness, conclusions without individualization: the new epistemology of forensic identification', *Law, Probability and Risk*, vol. 8, pp. 233-255.

Crane, A 1999, 'Does the amount of handwriting on a cheque constitute a 'reasonable amount of sample'?, *Canadian Society Forensic Science Journal*, vol.32, no. 1, pp. 39-45.

Cummaudo, M, Guerzoni, M, Marasciuolo, L, Gibelli, D, Cigada, A, Obertová, Z, Ratnayake, M, Poppa, P, Gabriel, P, Ritz-Timme, S, Cattaneo, C 2013, 'Pitfalls at the root of facial assessment on photographs: a quantitative study of accuracy in positioning of facial landmarks', *International Journal of legal Medicine*, vol. 127, pp.699-706.

Edmond, G 2008, 'Specialised knowledge, the exclusionary discretions and reliability: Reassessing incriminating expert opinion evidence', *University of New South Wales Law Journal*, vol.31, no. 1, pp. 1-55.

Edmond, G, Biber, K, Kemp, R, Porter, G 2009, 'Law's looking glass: expert identification evidence derived from photographic and video images', *Current Issues in Criminal Justice*, vol. 20, no. 3, pp. 337-376.

Edmond, G 2010, ‘Impartiality, efficiency or reliability? A critical response to expert evidence law and procedure in Australia’, *Australian Journal of Forensic Science*, vol. 42, no. 2, pp. 83-99.

Farkas, LG 1994, *Anthropometry of the Head and Face*. Raven Press, New York.

Gibelli, D, Brandone, A, Andreola, S, Porta, D, Giudici, E, Grande M, Cattaneo, C 2010, ‘Macroscopic, microscopic and chemical assessment of gunshot lesions on decomposed pig skin’, *Journal of Forensic Science*, vol. 55, no.4, pp.1092-1097.

Gordon, CC, Bradtmiller, B, Churchill, T, Clauser, CE, McConville, JT, Tebbetts, IO, Walker, RA 1988, ‘Anthropometric Survey of US Army Personnel: Methods and Summary Statistics’, Technical Report NATICK/TR-89/044, United States Army Natick Research, Development and Engineering Center, Natick, Massachusetts.

Goldstein, AJ, Harmon, LD, Lesk, AB 1971, ‘Identification of human faces’, *Proceedings of the IEEE*, vol.59, no. 5, pp.748-760.

Henneberg, M 2007, ‘Facial mapping, body mapping and the duties of an expert witness’, http://www.lawlink.nsw.gov.au/lawlink/pdo/ll_pdo.nsf/pages/PDO_facialmapping.

Henneberg, M 2008, ‘Expert witness in a courtroom: The Australian experience’ In: Oxenham, M (eds): *Forensic approaches to death disaster and abuse*. Australian Academic Press, Queensland. pp. 307-319.

Henneberg, M 2010, ‘The Illusive concept of human variation: Thirty years of teaching biological anthropology on four continents’ In Štrkalj G (ed.) *Teaching human variation*. Nova Science, New York, pp. 33-43.

Jafri, R, Arabnia, H 2009, ‘A survey of face recognition techniques’, *Journal of Information Processing Systems.* vol. 5, no. 2, pp.41-68.

Jain, AK, Hong, L, Pankanti, S, Bolle, R 1997, ‘An identity authentication system using fingerprints’. *Proceedings of the IEEE,* vol. 85, pp.1365-1388.

Jain, AK, Prabhakar, S, Pankanti, S 2002, ‘On the similarity of identical twin fingerprints’, *Pattern recognition,* vol. 35, 2653-2663.

Jayaprakash, PT 2013, ‘Practical relevance of pattern uniqueness in forensic science’, *Forensic Science International,* vol.231. pp. 1-16.

Jeffreys, AJ, Wilson, V, Thein, SL 1985, ‘Individual ‘specific’ fingerprints of humans DNA’, *Nature*, vol. 316, pp. 76-79.

Jobling, MA, Gill, P 2004, ‘Encoded evidence: DNA in forensic analysis’, *Nature Reviews Genetics,* vol.5, no.10, pp. 739-751.

Klare, B, Jain, AK 2010, ‘On a taxonomy of facial features’ *Biometrics: Theory Applications and Systems (BTAS),* 2010 Fourth IEEE International Conference on (pp. 1-8). IEEE.

Kleinberg, K, Vanezis, P, Burton, AM 2007, ‘Failure of anthropometry as a facial identification technique using high quality photographs’, *Journal of Forensic Science.* vol. 52, no. 4, pp. 779-783.

Mahaja, A, Bartra, A, Khurana, B, Kaur, J 2012, ‘Role of bitemark analysis in identification of a person’, *Global Journal of Medicine and Public Health.* vol. 1, no. 1, pp. 56-59.

McDonald, P 1993,*Tire imprinting evidence,* CRC Press, Florida.

Meijerman, L, Sholl, S, De Conti, F, Giacon, M, Van Der Lugt, C, Drusini, A, Vanezis, P, Maat, G 2004,‘Exploratory study on classification and individualisation of earprints’, *Forensic Science International,* vol. 140, no. 1, pp.91-99.

Mishra, G, Ranganathan, K, Saraswathi, T 2009, ‘Study of lip prints’, *Journal of Forensic Dental Sciences,* vol. 1, pp.28.

Oatess, RT 2000, ‘Elbow print identification’, *Journal of Forensic Identification.* vol.50, pp.132-137.

Page, M, Taylor, J, Blenkin, M 2011, ‘Uniqueness in the forensic identification sciences-fact or fiction?’, *Forensic Science International*. vol.206, pp. 12-18.

Reichs, KJ 1998, ‘Postmortem dismemberment: recovery, analysis and interpretation’, *Forensic Osteology: Advances in the identification of human remains*, pp. 353-388.

Ritz-Timme, S, Gabriel, P, Tutkuviene, J, Poppa, P, Obertová, Z, Gibelli, D, De Angelis, D, Ratnayake, M, Rizgeliene, R, Barkus, A, Cattaneo, C 2011, ‘Metric and morphological assessment of facial features: A study of three European populations’, *Forensic Science International*, vol. 207, pp.239e1-239e8.

Rodriguez-Celis, EM, Gornushkin, IB, Heitmann, UM, Almirall, JR, Smith, BW, Winefordner, JD, Omenetto, N 2008, ‘Laser induced breakdown spectroscopy as a tool for discrimination of glass for forensic application’, *Analytical and Bioanalytical Chemistry*, vol. 391, pp. 1961-1968.

Rosing, F 2006, *Identification von Personen auf Bildern*. Verlag C.H Beck, München.

Scoleri, T, Lucas, T, Henneberg, M 2014, ‘Effect of garments on photoanthropometry of body parts: Application of stature estimation’, *Forensic Science International.* vol. 237, pp. 1-12.

Scoleri, T, Henneberg, M 2012, ‘View-independent prediction of body dimensions in crowded environments’, In *International Conference on Digital Image Computing Techniques and Applications* (2012: Fremantle) DICTA.

Shi, K, Samal, A, Marx, D 2006, ‘How effective are landmarks and their geometry for face recognition?’, *Computer Vision and Image Understand*, vol. 102, pp.117-133.

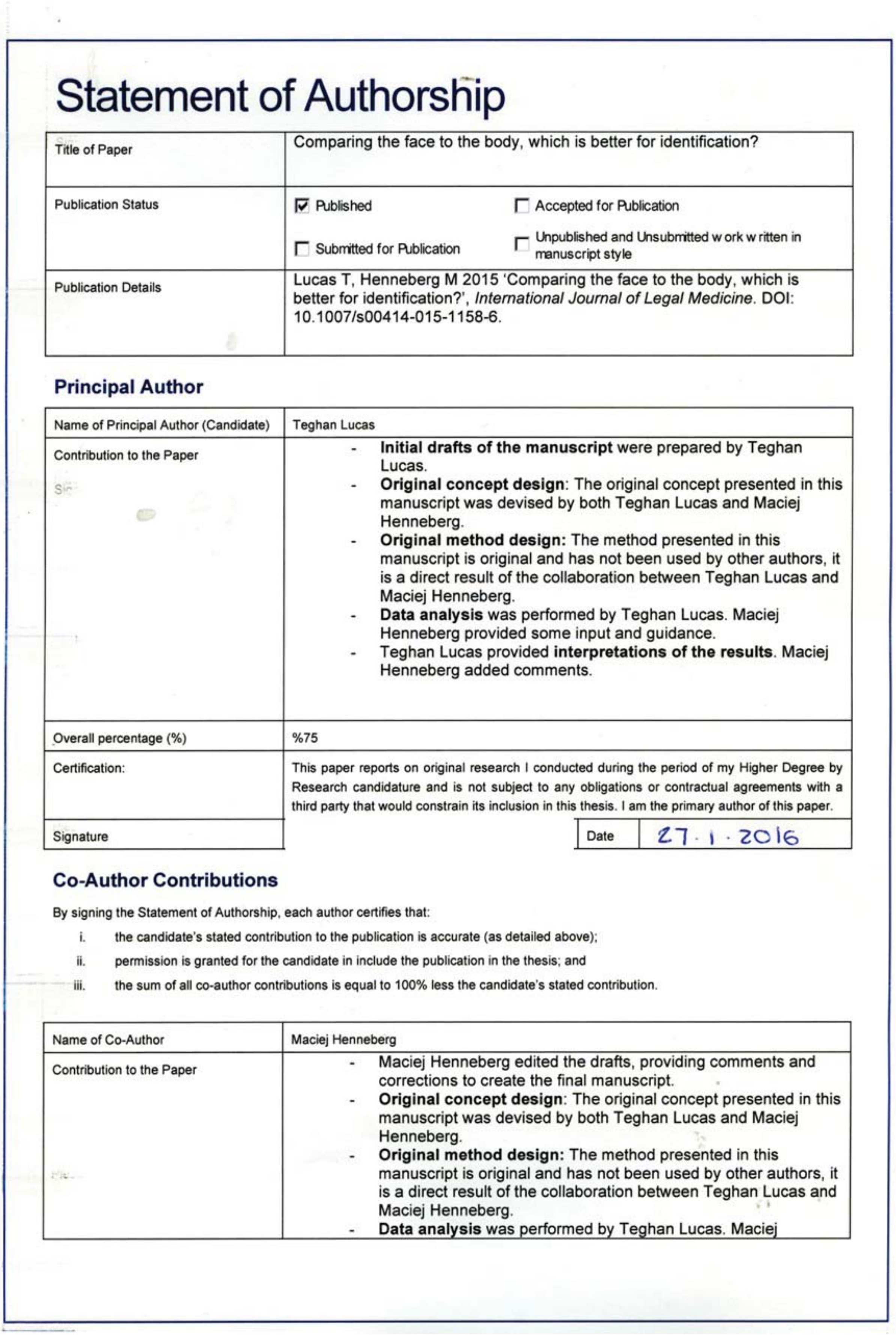
Spaun, NA 2007, ‘Forensic biometrics from images and video at the Federal Bureau of investigation’, In *Biometrics: Theory Applications and Systems (BTAS),*pp. 1-3.

Spaun, N 2009, ‘Facial comparisons by subject matter experts: Their role in biometrics and their training’, in *International Conference on Advances in Biometrics*, pp. 161-168.

Turk, M, Pentland, A 1991, ‘Eigenfaces for recognition’, *Cognitive Neuroscience*, vol.3, pp.72-86.

Springer, E 1995, ‘Toolmark examinations – A review of its development in the literature’, *Journal of Forensic Sciences*, vol. 40, no. 6, pp. 964-968.

Statement of Authorship Manuscript 3:



Statement of Authorship

Title of Paper	Comparing the face to the body, which is better for identification?	
Publication Status	☑ Published ☐ Submitted for Publication	☐ Accepted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Lucas T, Henneberg M 2015 'Comparing the face to the body, which is better for identification?', *International Journal of Legal Medicine*. DOI: 10.1007/s00414-015-1158-6.	

Principal Author

Name of Principal Author (Candidate)	Teghan Lucas
Contribution to the Paper	- **Initial drafts of the manuscript** were prepared by Teghan Lucas. - **Original concept design**: The original concept presented in this manuscript was devised by both Teghan Lucas and Maciej Henneberg. - **Original method design:** The method presented in this manuscript is original and has not been used by other authors, it is a direct result of the collaboration between Teghan Lucas and Maciej Henneberg. - **Data analysis** was performed by Teghan Lucas. Maciej Henneberg provided some input and guidance. - Teghan Lucas provided **interpretations of the results**. Maciej Henneberg added comments.
Overall percentage (%)	%75
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 27.1.2016

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Maciej Henneberg
Contribution to the Paper	- Maciej Henneberg edited the drafts, providing comments and corrections to create the final manuscript. - **Original concept design**: The original concept presented in this manuscript was devised by both Teghan Lucas and Maciej Henneberg. - **Original method design:** The method presented in this manuscript is original and has not been used by other authors, it is a direct result of the collaboration between Teghan Lucas and Maciej Henneberg. - **Data analysis** was performed by Teghan Lucas. Maciej

	Henneberg provided some input and guidance. - Maciej Henneberg added comments regarding the interpration of results performed by Teghan Lucas.		
Signature	[redacted]	Date	27.1.16

Please cut and paste additional co-author panels here as required.

3. Comparing the face to the body, which is better for identification?

Teghan Lucas[a*] and Maciej Henneberg[a]

[a] Biological and Comparative Anatomy Research Unit, Medical School, University of Adelaide, Adelaide 5005, Australia.

Submitted to International Journal of Legal medicine on the 4th of October 2014, accepted on 28th January 2015. **Published** on 10th February 2015.

[*]Corresponding author.

Teghan Lucas

Medical School North Building, Frome Road, The University of Adelaide, South

Australia 5005, Phone: +61 8 8313 6324. Email address: teghan.lucas@adelaide.edu.au

Context

When anthropometry was first introduced into forensic investigations, measurements of the head/face and body were used. However, over time, research has narrowed down to identify individuals only from facial measurements. Studies claim that the face is used as humans have evolved to identify others from the face (Barrett 2008), It is not covered by clothing and is the most recognised part of an individual as we look at the face during communications (Henneberg 2007). Surveillance images are becoming a more popular means of identification; thus criminals are covering their faces to avoid identification. Much research has concentrated on taking measurements from the images, therefore the issues surrounding 'recognition' as opposed to 'identification' are irrelevant. Therefore, the usefulness of body measurements for identification purposes needs further investigation.

The aims of this study were to establish whether or not measurements of the body were just as, if not, more useful than the measurements of the face. The same method of 'duplication' that was presented in Manuscript 2 'Are human faces unique? A metric approach to finding single individuals without duplicates in large samples' was used to investigate these aims.

As mentioned previously, much research has been focused on identification from the face (Bachi et al. 2014; Best-Rowden et al. 2014; Mukane, Hundiwale and Dere 2014; Jain, Klare and park 2012; Chen et al. 2010) the literature is constantly expanding. The usefulness of images of the human body for identification has been much less explored. To our knowledge, no other study investigated the usefulness of body measurements alone in isolating an individual from a large population. Furthermore, no study has calculated the probabilities of finding two or more individuals with the same set of body measurements, an issue which is extremely useful in court proceedings, similar to DNA (Jobling and Gill 2004) and fingerprint (Jain, Prabhakar and Pankanti 2002; Pankanti, Prabhaker and Jain 2002) studies. Current research showed that the body was better than the face for identification and was comparable with DNA and fingerprint studies.

This manuscript builds on the theory of the previous manuscript ‘Are human faces unique? A metric approach to finding single individuals without duplicates in large samples’ by establishing which part of the human body proves more useful for identification from measurements.

Some difficulties were involved with the presentation of findings. As manuscript 2 was being reviewed when this paper was being written, we could not use the term ‘singularity’ as it had not yet been published and presenting the term would have been too many ideas for one paper. This was resolved by presenting the inverse of ‘singularity’ which was individual cases, this avoided any reference to an unpublished paper and proved successful in portraying the results.

Abstract

As early as the 19th century, measurements of the face and body were used for forensic identification. It was believed that no two individuals had the exact same measurements. However, this was overtaken by fingerprint analysis because it was considered more reliable in court proceedings as the probabilities of finding matching individuals could be calculated. With the standardisation of photographs, identification primarily occurs from the face. With the ability to take measurements from photographs, why not use the body? The ANSUR database contains anthropometric measurements of 3982 individuals. Eight facial and eight body measurements were compared to investigate whether or not there is enough information on the body to use for identification. Measurements were compared by adding one measurement to the other(s) in a stepwise approach until there were no duplicate cases where two or more individuals share the same combination of measurements. Results consistently show that less body measurements are needed to find no duplicates when compared to the face. The larger the range of each of the measurements, the less chance there is of finding a duplicate. With the combination of eight body measurements it is possible to achieve a probability of finding a duplicate to the order of 10^{-20} or 1 in a quintillion. These results are comparable with fingerprint analysis. The body is more variable than the face and should be used in identification. An advantage to using the body is that larger dimensions are easier to locate on images and not affected by facial expression.

Key words: Physical anthropometry, forensics, duplication, ANSUR

Introduction

Humans have evolved to recognise faces (Barrett 2008). The face is used for recognition because it is not covered by clothing, it contains a large number of features, and it is the part of a person that we look at when communicating with them (Henneberg 2007). Thus, it is no mystery why the face is used in the forensic sciences for identification.

In the19th century Alphonse Bertillion (1886;1890), a French police officer created an identification system using anthropometric measurements of the face and body. Anthropometric measurements were recorded each time an individual was arrested and each time the results were compared to the previous measurements to establish if it was the same person, if the measurements taken matched previous measurements then they were considered to be the same individual and an identification was made. Thus it was assumed that no other individual had the exact same combination of measurements. Bertillion’s methods could only be used on repeat offenders as images of a person committing a crime were not available. Bertillion’s method was soon replaced by fingerprint analyses as criminals left their fingerprints at the crime scene which could then be compared to the suspect, thus no repeat in criminal activity was necessary for identification. Another advantage of fingerprint analyses is that the probability of finding two individuals with matching fingerprints was quantified, this allowed statements to be made in court proceedings pertaining to the reliability of using fingerprints as evidence (Jain, Prabhakar and Pankati 2002; Pankati, Prabhakar and Jain 2002). Bertillion also introduced the standardised photograph of a criminal’s face (mugshot). From this, the standardisation of facial photographs for identification has been incorporated into everyday life on passports, student ID cards and drivers licenses to name just a few. The face is a key area of study in the forensic sciences, the literature is constantly expanding (Bagchi et al. 2014; Best-Rowden et al. 2014; Mukane, Hundiwale and Dere 2014). Automated recognition systems are being developed to accurately identify and isolate someone by the face (Best-Rowden et al. 2014; Jain, Klare and Park 2012), thus reinforcing the belief that the face is the best part of the body to use for identification.

In forensic identification, one of the key questions is what are the chances that another person has the same features, could there be a duplicate of this individual? Ideally, this question would be answered by observing characteristics of every individual on the Earth at any one given time. Let's face it, the entire population of the Earth cannot be observed. Therefore, the question of finding a possible duplicate is most commonly answered using statistical probabilities based on research conducted on sample sizes smaller than the Earth's population. Fingerprint analysis is one of the better known areas that use probabilities to support forensic evidence (Jain, Prabhakar and Pankanti 2002; Pankanti, Prabhakar and Jain 2002). Often the probability of finding two individuals with identical fingerprints exceeds what the population of the Earth can provide ie. 10^{10}. In forensics, it is desirable to use characteristics objectively measurable for which probabilities can be calculated. The lower the probability of finding matching characteristics the more reliable for identification the characteristics are considered to be. The basic principle behind these forensic techniques is the more information is available about an individual, the less chance there is of finding another individual matching. For example, if facial characteristics are considered quite a few people with a combination of blue eyes, Darwin's tubercle, detached earlobe and thick lips can be found. However, the chance of finding an individual with all of the above characteristics plus thirty more traits is extremely low, if not impossible. Thus the key to successful identification methods lies in the combination of largely independent characters.

With the proliferating use of Closed Circuit Television (CCTV) it is no longer true that criminals do not leave their anatomical characteristics behind. Image based identification has become increasingly popular, despite the fact that the images they produce are of poor quality (Chen et al. 2010). Expert witnesses in biological anthropology have been using descriptive methods as early as the 1900s in an attempt to identify individuals (Knussman 1983, 1988; Rosing 2006, 2013). Descriptive methods have been criticised on the basis that they are unreliable (Edmond 2008, 2010; Edmond et al, 2009) and images are open to interpretation by the person analysing them (Biber 2009). In an attempt to make facial identification and recognition more reliable, research is concentrating on methods which take metric measurements from images (Scoleri and Henneberg 2012; Scoleri, Lucas and Henneberg 2014). However, it is important to note that these methods are still being developed and are often dependent on optimal lighting, pose and high image quality (Zhao et al. 2003). Facial features are hard to detect on low quality images as often fine details cannot be seen (Chen et al. 2010) or they are purposely covered to prevent identification (Henneberg 2007).

It seems as though some of Bertillion’s message was lost, he originally took anthropometric measurements of the body, so why do we use only the face for identification, why not the body?

As mentioned earlier, the face is primarily used for identification as it is not covered by clothing. This is hardly the case when committing a crime as many criminals cover their face to avoid identification (Henneberg 2007). The argument of not using the body for identification because of clothing, implies that clothing somehow obscures body characteristics such as shape. It has been suggested that the colour (Frith and Gleeson 2004) size (Fan, Yu and Hunter 2004) and pattern (Meekins 2006) of clothing can alter an individual’s perception of body shape. However, Lucas, Kumaratilake and

Henneberg (2014) investigated these claims and found that description of an individual's body shape was little influenced by different types of clothing. A possible reason for this is that gravity acts on clothing in such a way that it hangs off the body, showing the overall shape (Henneberg 2007). Many of the measurements of the body can still be obtained from clothed individuals since many anthropometric points can be located on the outline/silhouette of an individual. This is different from facial characteristics many of which are located in the middle of the face and thus poorly visible on pictures where only silhouette can be seen. Although the body is covered by clothing, any movement changing position of the jointed body parts can only increase the chances of seeing the outline of an individual. Larsen et al. (2008) found that anatomical points could be located better when the joints were not extended.

The fact that we primarily recognise individuals based on facial features would suggest that we are rather exceptional at it, we are not. Although humans have an exceptional ability to recognise familiar faces (even in low quality images), we are no good at recognising unfamiliar faces (Hancock, Bruce and Burton 2000; Henderson, Bruce and Burton 2001). This is rather concerning considering the forensic sciences and other fields rely on the recognition of unfamiliar faces in order to make an identification. Recognition rates are also affected by facial expression, something which can be easily changed many times throughout the day. Our natural inability to recognise unfamiliar faces paired with the difficulties associated with facial expression is one of the reasons why descriptive methods are criticised. If we are shifting from descriptive to metric methods it is no longer relevant which part of an individual we primarily use for recognition, in short, metrics can be applied anywhere.

There has been little research conducted on identifying individuals based on body measurements. Most of the studies which mention the body are doing so in reference to gait analysis (Bouchrika et al. 2011; Larsen, Simon andLynnerup 2008). Some studies have proved successful in taking measurements of the body from images (Scoleri and Henneberg 2012; BenAbdelaker and Davis 2006; BenAbdelaker and Yacoob 2008). Others have reconstructed one measurement (ie. height) from another (ie. upper limb length) (Scoleri, Lucas and Henneberg 2014). However, these studies do not use multiple body measurements nor discuss the body's usefulness in identification. Currently, body measurements are considered a secondary source of information, meaning they are not to be used on their own for identification but can be used with more 'useful' sources such as facial recognition, fingerprints and DNA (Jain, Dass and Nandakumar 2004). The reason for this is because body characteristics are *a priori* considered to contain inadequate information to isolate an individual from a population ie. Identify them.

Therefore, this study aims to establish whether or not body measurements can be just as, if not more useful than face measurements.

Materials

Sample

The U.S Army Anthropometric Survery (ANSUR) was conducted in 1988. Although the sample was obtained over 20 years ago, the data are still valid for this study as they characterise normally varying large samples of modern humans. The sample consists of 1774 men and 2208 women aged 17-51 years. The ANSUR database contains information about the ancestry of all participants, ANSUR refers to these groups as 'ethnic groups', 66.1% of the men and 51.6% of women are White; 25.8% of men and 41.8% of women are Black while only 3.8% of men and 2.6% of women are

Hispanic. The rest of the sample is divided over other groups consisting of Asians, Pacific Islanders, Native Americans etc. Details of this study are described in (ANSUR 1988). For all analyses, males and females are combined to increase the sample size. This is justified as only about 25% of morphological variation occurs between sexes (Henneberg 1990; 2010). It is reported that upwards of 95% of the variation of a trait occurs within populations instead of between them (Cavalli-Sforza and Bodmer 1971). Therefore this study did not separate participants by the ‘ethnic groups’ reported by ANSUR. All measurements are reported to the nearest millimetre. All measurements included in this study are considered accurate. Every care was taken to ensure human rights by the ANSUR measurement team, including; debriefing participants on procedures, confidentiality and obtaining written consent.

Anthropometric measurements

A team of 22 individuals were responsible for taking the anthropometric measurements. Each measurer was trained for four weeks before taking any measurements of the subjects in this study. Measurers were assigned specific dimensions, they were taught by trained anthropometrists how to locate the landmarks and measure each dimension. The reported reliability coefficients for anthropometric dimensions (how much of the variation between participants in the measured sample is free from measurement error) range from 90.3% to 99.8% for all measurements (Gordon et al. 2013).

The following body measurements were used (Fig.1): *Calf circumference* – The maximum horizontal circumference on the right calf. *Chest circumference* – The maximum circumference of the chest at the fullest part of the breast. *Forearm* *circumference* - The circumference at the elbow crease of the forearm. *Hip breadth* – The horizontal distance between the lateral buttock landmarks on the sides of the hips. *Iliocristale height* – The vertical distance between the floor and the iliocristale landmark on the right side of the pelvis. *Radiale – stylion length* - The distance between the radiale landmark on the right elbow and the stylion landmark on the right wrist. *Stature* – The vertical distance from the floor to the top of the head (vertex). *Waist circumference* – The horizontal circumference of the waist at the level of the centre of the navel (omphalion). These measurements were chosen from the ANSUR database at random.

The following face/head measurements were used (Figure 1): *Bitragion submandibular arc* – The surface distance between the right and left tragion across the submandibular landmark at the juncture of the jaw and the neck. *Bizygomatic breadth* – The maximum horizontal breadth of the face between the zygomatic arches. *Ear length* – The length of the right ear from its highest to lowest points on a line parallel to the long axis of the ear. *Ear protrusion* – The horizontal distance between the mastoid process and the outside edge of the right ear at its most lateral point (ear point). *Head breadth* – The maximum horizontal breadth of the head above the ears. *Head circumference* – the maximum circumference of the head above the supraorbital ridges and ears. *Head Length* – the distance from the glabella landmark between the brow ridges to opisthocranion. *Interpupillary distance* – the distance between the centres of the right and left pupils. These measurements were chosen as they are not influenced by facial expression.

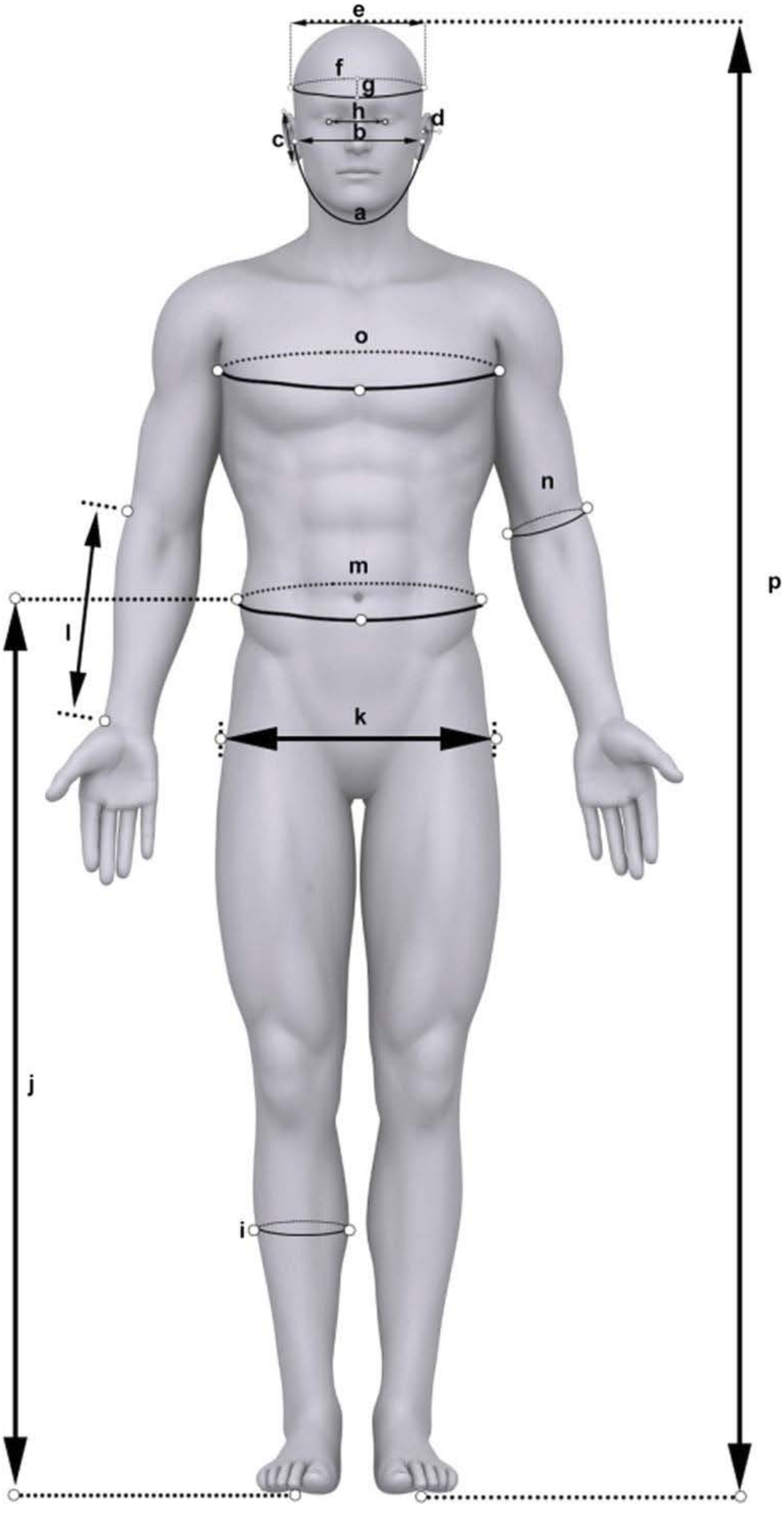


Figure 1: Anterior view of the entire body showing: (a) bitragion submandibular arc (b) bizygomatic breadth (c) Ear length (d) Ear protrusion (e) Head breadth (f) head circumference (g) head length (h) interpupillary distance (i) calf circumference (j) iliocristale height (k) hip breadth (l) radiale-stylion length (m) waist circumference (n) forearm circumference (o) chest circumference (p) stature

Method

Performance of anthropometric measurements of the head/face in finding no duplicate individuals in the sample is compared to those taken from the body. Each set of measurements (head/face and body) is searched for combinations of duplicate measurements among individuals. A duplicate measurement is any measurement of an individual's trait ie. stature which matches numerically that of another individual(s). In this paper, we used measurements in millimetres to demonstrate how duplicates can be found. We realise that anthropometric dimensions of same individuals can vary to some extent due to a number of factors. If those errors are known, the numerical values can be modified.

The number of individuals who are not duplicates of each other indicates how well a combination of measurements can individualise a person. In this analysis traits are added to others in a stepwise approach until no duplicates are found and the number of individual cases reaches that of the sample size (N=3982). For example, in the sample of 3982 individuals, for the first measurement analysed bitragion submandibular arc the number of individuals without duplicates was 112, when head circumference was added the number of individuals without duplicates increased to 2,477, when head length was added the number of individuals without duplicates increased to 3,808 . This was continued by adding head breadth and bizygomatic breadth until no duplicates were found and all individuals were considered as individual cases. The list of traits was changed throughout the analyses to establish if the order of traits had any effect.

As long as the traits are not perfectly correlated with each other (r <1.00) the addition of each trait adds more information and reduces the number of duplicates. The addition of traits to previous traits continues in both sets of measurements until no duplicates are found and all cases are individual. The least number of traits combined needed to have the entire sample of all individual cases ie. no duplicates is most favourable for identification purposes.

In order to determine the probability of finding two matching individuals based on face/head and body characteristics the following calculations were made. The number of units (range) is calculated for all measurements of the face/head and the body, the range is the difference between the minimum and maximum size of the trait expressed in the units of measurement (Table 1). Due to the intercorrelation of traits, the possible number of independent combinations is limited. Using a stepwise approach, the adjusted multiple regression R squared values were calculated for each set of traits (face and body). The R squared value was subtracted from 1 to determine the portion of the trait's variance which is not correlated with others. This allowed traits to be combined randomly. The number of combinations of traits which are free of intercorrelations is $1-R^2$ multiplied by the number of possible units. Reciprocal of the number of combinations tells us the likelihood that another individual shares the exact same measurements of the two combined traits. For example. If two traits are used, the first trait with a range of 614 mm and the second with a range of 570 mm the number of possible combinations of units for those two traits would be 614 x 570 = 349980 units were they not correlated. If the coefficient of determination (R^2) of these two traits is 0.284 then $1-R^2$ would be 0.716. The number of combinations becomes 0.716* 349980 = 250586 (=k). Then the probability (p) of finding duplicates based on these traits is p= 1/k = 1/250586 = 0.000004 or 1 out of a quarter million. This is repeated in a stepwise approach until no traits are left. IBM SPSS statistics 20 was used for all statistical analyses. Microsoft Excel was used for graph generation.

Results

Table 1 shows the head/face traits used and their ranges in millimetres. Table 2 shows the body traits used and their ranges in millimetres. The order of both lists is alphabetical and will change throughout analyses as indicated.

Table 1: List of traits on the face and their range (mm) for the sample of 3982 individuals

Facial Measurements	
Trait	Range (mm)
Bitragion submandibular arc	138
Bizygomatic Breadth	44
Ear Length	35
Ear Protrusion	27
Head Breadth	47
Head Circumference	127
Head Length	62
Interpupillary Distance	26

Table 2: List of traits on the body and their range (mm) for the sample of 3982 individuals.

Body Measurements	
Trait	Range (mm)
Calf Circumference	185
Chest Circumference	570
Forearm circumference	162
Hip Breadth	150
Iliocristale Height	502
Radiale-Stylion Length	168
Stature	614
Waist circumference	554

The order of traits has been changed from alphabetical to smallest to largest based on the range of the traits (Table 1). The order of traits has been changed to establish any differences that the range of each trait may have on the results when smaller ranged traits are used first. Figure 2 shows the percentage of individual cases for the face and the body. The number of individual body cases increases by 71.3% when the second body trait (forearm circumference) is added to the first (hip breadth), this is the largest increase shown. It takes 5 body traits as opposed to 7 facial traits to find all persons to be individual cases (no duplicates) in the sample.

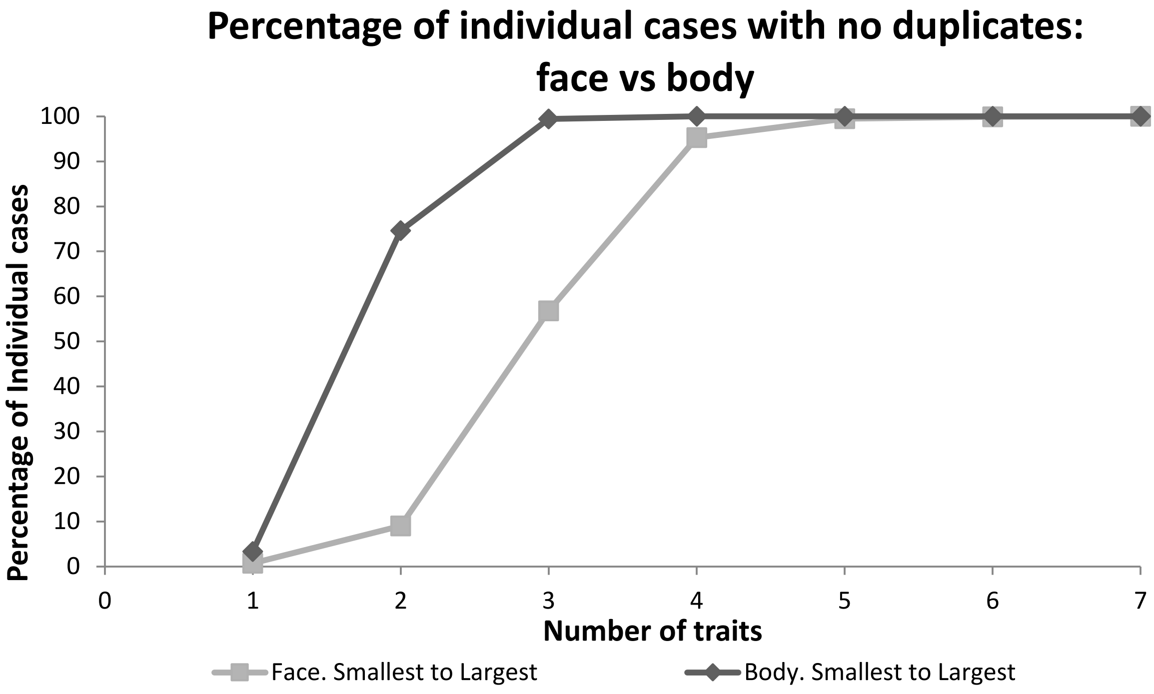


Figure 2: The percentage of individual cases for the face and the body (N=3982) when the list of traits is ordered smallest to largest based on their range.

The order of traits for the body has been changed to largest to smallest based on the range (Table 2). Figure 3 shows the percentage and number of individual cases with no duplicates for the body and the face. In both the face and the body there is only a small increase in the number of individual cases when only one trait (stature or bitragion submandibular arc) is analysed. Once the second body trait (chest circumference) is added an increase of 86.2% of individual cases is observed. Only three traits for the body are needed until the entire sample (N=3982) consists of individual cases (no duplicates) ; whereas five facial traits are needed to find no duplicates within the sample. When using faces, there is also less of an increase in the number of individual cases per added trait when compared to the body.

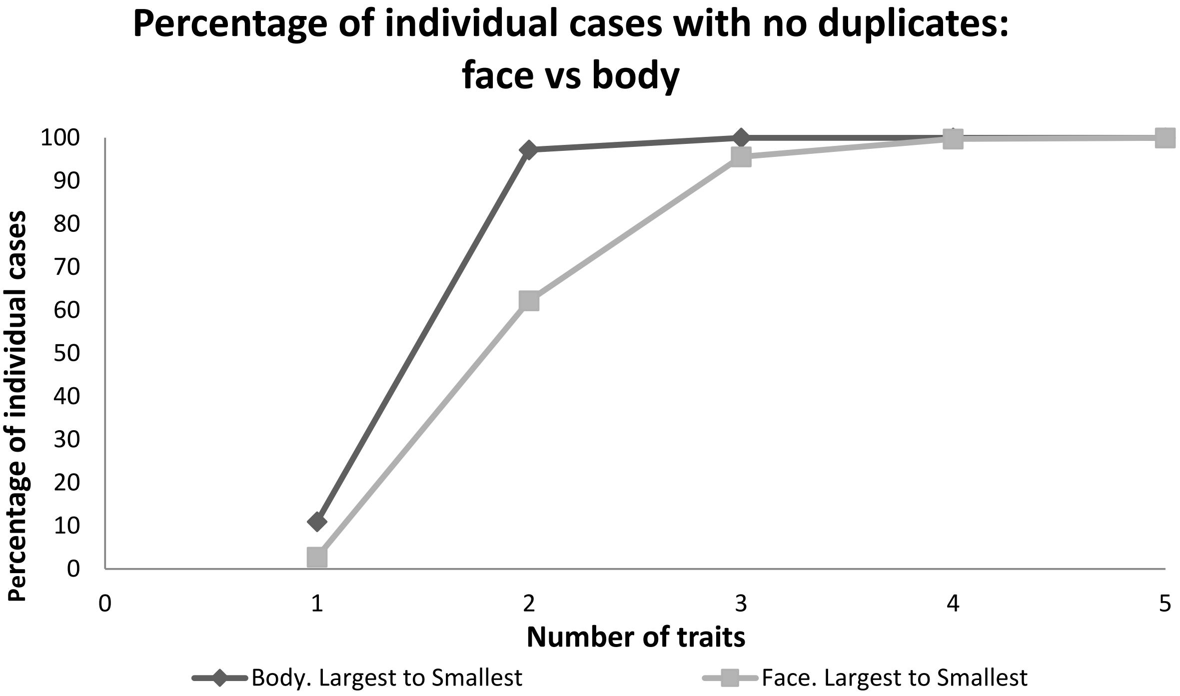


Figure 3: The percentage of individual cases for the face and the body (N=3982) when the traits are ordered largest to smallest.

Figure 4 shows the probability of finding another individual which matches one within the ANSUR sample when 1 to 8 traits are used. The more traits used, the less chance there is of finding an exact duplicate. The order of the traits used is largest to smallest. The best result is seen with the traits of the body, it is 10^{-20} or 1 in a quintillion. The face produced probabilities of 10^{-14} which are close to 1 in a trillion.

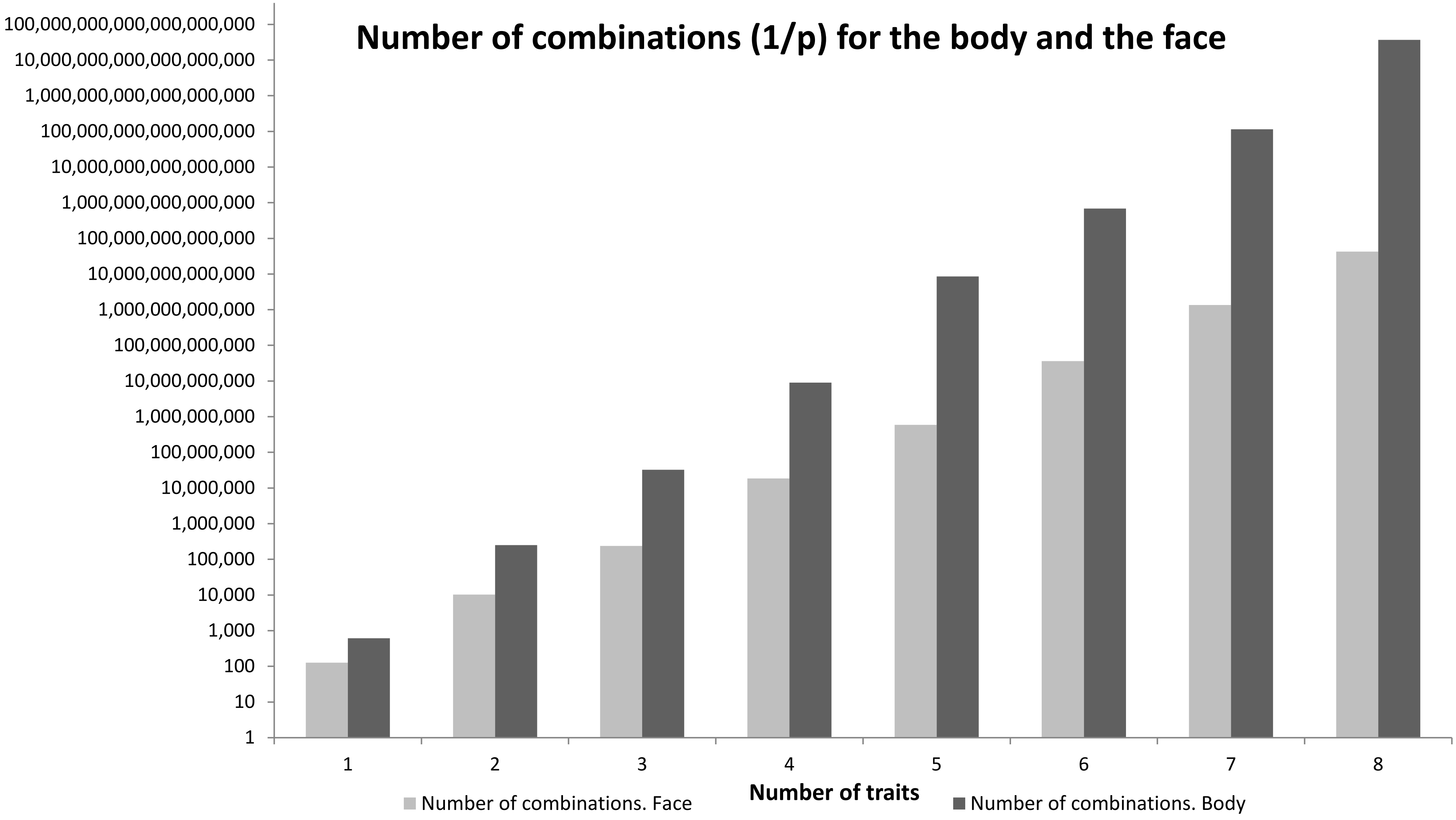


Figure 4: The number of combinations (1/p) for the body and the face.

Discussion

The forensic identification relies on the combination of traits, the more information is included, the less chance there is of finding an exact duplicate of an individual. This is well accepted in relation to the face. Using metric traits of the body is even better. The larger the number of traits and their respective range, the faster the number of individual cases increases and thus the less traits are needed to find no duplicates. This is shown consistently throughout the analyses. Even when body traits which have the smallest range are compared to facial traits with the largest range, they still produce better results. Therefore, body traits should always be chosen before the face. This has particular importance for forensic cases, especially those which use low quality CCTV images as often smaller details such as those in the face are not visible. This also increases the objectivity of analyses because both, investigators and the court are not influenced by the well-established preconceptions concerning faces.

Facial expression influences the measurement and perception of traits (Wang, Tan and Jain 2003) therefore in this study facial traits little influenced by facial expression were chosen. Body measurements were chosen at random and still produced better results when compared to the ideal conditions for the face (traits not affected by expression).

The probability of finding another individual with the exact same eight body characteristics is 1 in a quintillion (Figure 4). This result equals or exceeds some fingerprint studies (Pankati, Prabhakar and Jain 2002). The fingerprint studies which show better results include more characters, therefore if more body characters were added to the list then the results will be comparable with, if not exceed the best fingerprint study. The results of this study are also comparable with results from DNA studies, the longer the fragment of DNA the more information can be extracted and therefore the lower the probability of finding another individual with the same sequence (Jobling and Gill 2004). This study demonstrates the basic principles of forensics sciences ie. the more information analysed, the less chance there is of finding two individuals who share the exact same traits.

This study has wider implications which can add to the knowledge of human variation. Choosing not to separate the participants by ‘ethnic’ groups or sex produced results that show each individual varies so much that they can be distinguished from all others just on their anthropometric measurements. In cases where evidence is given regarding ‘ethnic’ groups or sex, this is considered descriptive and can often by misinterpreted if the images are of low quality. In forensic anthropology, ‘race’ or ‘ethnicity’ is often assigned based on anthropometric measurements (Farkas, Katic and Forrest 2005; Zhuang et al. 2010), this study validates claims that more variation occurs within populations rather than between them.

One of the arguments for not using body measurements for identification is that they lack permanence which is essential when comparing people over a long time period (BenAbdelaker and Yacoob 2008) which is sometimes the case for forensics. If body measurements which use vertical and horizontal distances between skeletal points (not circumferences) are used then these anthropometric measurements are permanent since an adult skeleton does not grow and identification of stable skeletal points is little influenced by overlying soft tissues. With extremely advanced age there may be some reduction of skeletal dimensions, but senile people are rarely subjects of forensic identifications. Therefore, these measurements would be permanent through time and body measurements can be used as accurately as fingerprints.

This paper is just an example of what can be achieved with a small number of metric traits of the body, it is in no way a guideline for which body traits should be used. Any body traits can be used and it is hypothesised that different traits will produce similar results. In identification from 2D images, circumferences cannot accurately be measured, therefore when using this method only lengths and widths of visible body parts should be measured. Each case is different and the number of measureable characteristics will change, this study provides an example of what can be achieved with a number of measurements and a large sample. However, circumferences are useful to demonstrate the amount of human variability in a large sample and may be useful for forensic identification with the incorporation of 3D whole body scanners.

To further demonstrate the usefulness of body measurements, a secondary sample of 1262 Australian women measured in 2002 (Henneberg and Veitch 2003; 2005) was searched for duplicates. Only 3 body traits (height, suprasternal height and dactylion height) were needed to find no duplicates. The method proposed in this paper can be applied to any sample. Humans are variable through time and geographic space. Therefore specific forensic cases may require databases specific for their circumstances. This may make construction of anthropometric databases specific for particular population at a particular time a worthwhile exercise.

References

ANSUR database, 1988, Clauser CE, Tebbetts IO, Bradtmiller B, McConville JT, Gordon CC Measurer's Handbook: US Army Anthropometric Survey 1987-1988, Technical Report NATICK/TR-88/043. Gordon CC, Bradtmiller B, Churchill T, Clauser CE, McConville JT, Tebbetts IO, Walker RA *1988.* Anthropometric Survey of US Army Personnel: Methods and Summary

Statistics, Technical Report NATICK/TR-89/044, United States Army Natick Research, Development and Engineering Center, Natick, Massachusetts.

Bagchi P, Bhattacharjee, D, Nasipuri, M, Basu, DK 2014, *Registration of three dimensional human face images across pose and their applications in digital forensic*. Springer International Publishing, Switzerland.

Barrett, HC 2008, ‘Evolved cognitive mechanisms of human behaviour’. In: Crawford C, Krebs D, editors. *Foundations of evolutionary psychology: ideas, issues, applications and findings*. (2nd ed.) Erlbaum Associates, New Jersey, pp. 173-190.

BenAbdelkader, C, Davis L 2006 ‘Estimation of anthropomeasures from a single calibrated camera’, in: *Proceedings of International Conference on Automatic Face and Gesture Recognition*, Southhampton pp. 499–504.

BenAbdelkader, C, Yacoob, Y, 2008 ‘Statistical body height estimation from a single image’, in: *Proceedings of International Conference on Automatic Face and Gesture Recognition*, Amsterdam, pp. 1–7.

Bertillon, A 1886, *Identification anthropometriques. Methode nouvelle de determination de L’iidentite individuelle*, Masson, Paris.

Bertillion, A 1890, *La Photographie judiciaure, avec un, appendice sur la classification et l’identification anthropometrique*, Gauthier-Villars, Paris.

Best-Rowden, L, Han, H, Otto, C, Klare, B.F, Jain, A.K 2014. ‘Unconstrained face recognition: Identifying a person of interest from a media collection’, *Information Forensics and Security, IEEE Transactions on*, *9*(12), pp.2144-2157.

Biber, K 2009, ‘Visual jurisprudence: the dangers of photographic identification evidence’, *Criminal Justice Matters*, vol.78, pp. 35-37.

Bouchrika, I, Goffredo, M, Carter, J, Nixon, M 2011, ‘On using gait in forensic biometrics’, *Journal of Forensic Sciences*, vol. 56, pp. 882-889.

Cavalli-Sforza, LL, Bodmer, WF 1971, *The genetics of human population*. WH Freeman and company, San Francisco.

Chen, S, Mau, S, Harandi, M, Sanderson, C, Bigdeli, A, Lovell, B 2010, ‘Face recognition from still images to video sequences; A local-feature-based framework’. *Journal of Image and Video Processing*, vol. 2011, pp.1-14.

Edmond, G 2008, ‘Specialised knowledge, the exclusionary discretions and reliability: Reassessing incriminating expert opinion evidence’. *University of New South Wales Law Journal*, vol. 31, pp.1-55.

Edmond, G 2010, ‘Impartiality, efficiency or reliability? A critical response to expert evidence law and procedure in Australia’, *Australian Journal of Forensic Science*, vol. 42, pp. 83-99.

Edmond, G, Biber, K, Kemp, R, Porter, G 2009, ‘Law's looking glass: expert identification evidence derived from photographic and video images’. *Current Issues in Criminal Justice*, vol. 20, pp. 337-376.

Fan, J, Yu, W, Hunter, L 2004, *Clothing appearance and fit: science and technology*, Woodhead Publishing, Cambridge.

Farkas, LG, Katic, MJ, Forrest, CR 2005, ‘International anthropometric study of facial morphology in various ethnic groups/races’, *Journal of Craniofacial Surgery*, vol. 16, pp. 615-646.

Frith, H, Gleeson, K 2004, ‘Clothing and embodiment: men managing body image and appearance’, *Psychology of Men and Masculinity*, vol. 5, pp.40-48.

Gordon, C, Claire, C, Blackwel,l CL, Bradtmiller, B, Parham, JL, Hotzman, J, Paquette, SP, Corner, BD, Hodge, BM 2010, *Anthropometric Survey of US Marine Corps Personnel: Methods and Summary Statistics*. No. NATICK/TR-13/018. Army Natick Soldier Research Development and Engineering Center MA, 2013.

Hancock, P, Bruce, V, Burton, M 2000,’Recognition of unfamiliar faces’, Trends in Cognitive Science, vol. 4, pp. 330-337.

Henderson, Z, Bruce, V, Burton, M 2001, ‘Matching the faces of robbers captured on video’, *Application of Cognitive Psychology*, vol.15, pp. 445-464.

Henneberg, M 2007, ‘Facial mapping, body mapping and the duties of an expert witness’.http://www.lawlink.nsw.gov.au/lawlink/pdo/ll_pdo.nsf/pages/PDO_facialmapping.

Henneberg, M 1990, ‘Brain size/body weight variability in Homo sapiens: consequences for interpreting hominid evolution’, *HOMO*, vol. 39, pp.121-130.

Henneberg, M 2010, The Illusive concept of human variation: Thirty years of teaching biological anthropology on four continents. In: Štrkalj, G. (eds) *Teaching human variation*. New York: Nova Science, p 33-43.

Henneberg, M, Veitch, D 2003, ‘National size and shape survey of Australia’, Kinanthreport. 16:34–39.

Henneberg, M, Veitch, D 2005, ‘Is obesity as measured by body mass index, and waist circumference in adult Australian women 2002 just a result of the lifestyle?’, *Journal of Human Ecology (Special Issue)*, vol. 13, pp. 85–89.

Jain, AK, Dass, SC, Nandakumar K, 2004, ‘Can soft biometric traits assist user recognition?’ *Defense and security*, International Society for optics and photonics.

Jain, AK, Klare, B, Park U 2012, ‘Face matching and retrieval in forensic applications’, *Multimedia*, vol.19, pp.20-28.

Jain, AK, Prabhakar, S, Pankanti, S 2002, ‘On the similarity of identical twin fingerprints’, *Pattern Recognition*, vol. 35, pp.2653-2663.

Jobling, MA, Gill, P 2004, ‘Encoded evidence: DNA in forensic analysis’. *Nature Reviews Genetics*, vol. 5, pp. 739-751.

Knussmann, R 1983, *Die vergleichende morphologische Analyse als Identitatsnachweis*. Strafverteidiger 3:127-129.

Knussmann, R 1988, *Anthropologie*. Bd I/1. Gustav Fischer, Stuttgart.

Larsen, PK, Hansen, L, Simonsen, EB, Lynnerup N 2008, ‘Variability of bodily measures of normally dressed people using photomodeler pro 5’. *Journal of Forensic Sciences*, vol. 53, pp.1393-1399.

Larsen, PK, Simonsen, EB, Lynnerup, N 2008, ‘Gait analysis in forensic medicine’, *Journal of Forensic Sciences*, vol. 53, pp. 1149-1153.

Lucas, T, Kumaratilake, J, Henneberg, M 2014, ‘The extent to which garments affect the body shapes of males from faceless CCTV images’, *Journal of Biological and Clinical Anthropology*, vol. 71, pp. 259-274.

Meekins, F 2006, Optical illusion wear, US Patent, Patent no. US:7,107,621, B2.

Mukane, D, Hundiwale, SM, Dere, P 2014, ‘Emerging forensic face matching technology to apprehend criminals: a survey’, *International Journal of Advances in Engineering and Technology*, vol. 7, pp. 255-262.

Pankanti, S, Prabhakar, S, Jain AK 2002, ‘On the individuality of fingerprints’, *Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 1010-1025.

Rösing, F 2006, *Identification von Personen auf Bildern*. Verlag C.H Beck, München.

Rösing, F 2013, Morphologische Identifikation von Personen. Grundlagen, Merkmale, Häufigkeiten. In: Buck J, Krumbholz B Hrg: *Sachverständigenbeweis im Verkehrs- und Strafrecht*. 2nd edn. Nomos-Verlag, Baden-Baden, pp 287-440.

Scoleri, T, Henneberg, M 2012, ‘View-independent prediction of body dimensions in crowded environments’. *International Conference on Digital Image Computing Techniques and Applications (DICTA)*; 3-5 December 2012; Fremantle.

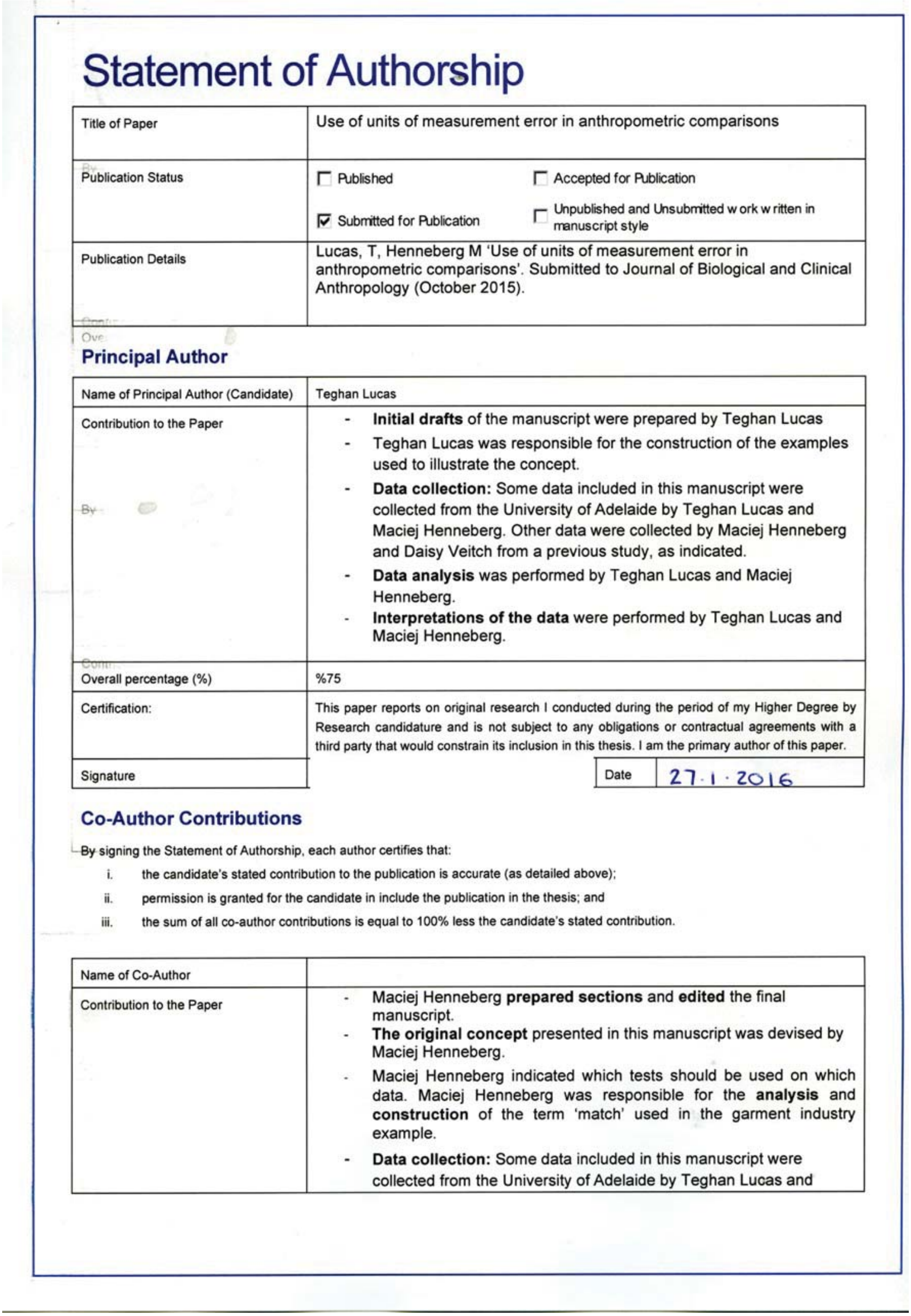
Scoleri, T, Lucas, T, Henneberg, M 2014, ‘Effect of garments on photoanthropometry of body parts: Application of stature estimation’, *Forensic Science International*, vol. 237, pp.1-12.

Wang, Y, Tan, T, Jain, AK 2003, Combining face and iris biometrics for identity verification. In *Audio and video based biometric person authentication*. Berlin: Springer, pp 805-813.

Zhao, W, Chellappa, R, Phillips, PJ, Rosenfeld, A 2003, ‘Face recognition: a literature survey’, *Computing Surveys*, vol. 35, pp. 399-458.

Zhuang, Z, Landsittel, D, Benson, S, Roberge, R, Shaffer, R 2010, ‘Facial anthropometric differences among gender, ethnicity and age groups’. *Annals of Occupational Hygiene*, vol. 1, pp. 1-12.

Statement of Authorship Manuscript 4:



Statement of Authorship

Title of Paper	Use of units of measurement error in anthropometric comparisons
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Lucas, T, Henneberg M 'Use of units of measurement error in anthropometric comparisons'. Submitted to Journal of Biological and Clinical Anthropology (October 2015).

Principal Author

Name of Principal Author (Candidate)	Teghan Lucas
Contribution to the Paper	- **Initial drafts** of the manuscript were prepared by Teghan Lucas - Teghan Lucas was responsible for the construction of the examples used to illustrate the concept. - **Data collection:** Some data included in this manuscript were collected from the University of Adelaide by Teghan Lucas and Maciej Henneberg. Other data were collected by Maciej Henneberg and Daisy Veitch from a previous study, as indicated. - **Data analysis** was performed by Teghan Lucas and Maciej Henneberg. - **Interpretations of the data** were performed by Teghan Lucas and Maciej Henneberg.
Overall percentage (%)	%75
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 27.1.2016

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	
Contribution to the Paper	- Maciej Henneberg **prepared sections** and **edited** the final manuscript. - **The original concept** presented in this manuscript was devised by Maciej Henneberg. - Maciej Henneberg indicated which tests should be used on which data. Maciej Henneberg was responsible for the **analysis** and **construction** of the term 'match' used in the garment industry example. - **Data collection:** Some data included in this manuscript were collected from the University of Adelaide by Teghan Lucas and

	Maciej Henneberg. Other data were collected by Maciej Henneberg and Daisy Veitch from a previous study, as indicated. - **Data analysis** was performed by Teghan Lucas and Maciej Henneberg. - **Interpretations of the data** were performed by Teghan Lucas and Maciej Henneberg.		
Signature		Date	27.1.16

Name of Co-Author			
Contribution to the Paper			
Signature		Date	

Please cut and paste additional co-author panels here as required.

4. Use of units of measurement error in anthropometric comparisons

Teghan Lucas[a*] and Maciej Henneberg[a,b]

[a] Biological and Comparative Anatomy Research Unit, Medical School, University of Adelaide, Adelaide 5005, Australia.

[b] Institute of Evolutionary Medicine, Universty of Zurich, Winterthurerstr. 190, 8057 Zurich, Switzerland

Submitted to Journal of Biological and Clinical Anthropology on the 16th November 2015.

[*]Corresponding author.

Teghan Lucas

Medical School North Building, Frome Road, The University of Adelaide, South

Australia 5005, Phone: +61 8 8313 6324. Email address: teghan.lucas@adelaide.edu.au

Context

Errors in anthropometry occur despite efforts to reduce them as much as possible (Stomfai et al. 2011; Perini et al. 2005; Ulijaszek and Kerr 1999). Errors cannot be avoided. One of the most widely used calculation for the measurement of error in anthropometry is the technical error of measurement (TEM) (Stomfai et al. 2011; Perini et al. 2005; Ulijaszek and Kerr 1999). Currently, the commonly used approach is to calculate errors and simply report them alongside the ‘reported values’. This introduces an ambiguity to the reported values. In large surveys, simply reporting errors is not practical for comparisons of participants or objects.

The aims of this study were to introduce a new method, which eliminates the ambiguity of analytical results based on measurements taken with random errors by replacing metric units of measurement with units of TEM. The advantages of using units of TEM were explored by applying the method to three different fields, which employ anthropometry. They are; forensic investigations, clothing industry and biological variation. Common statistical methods within these fields were used and the results are expressed in metric units and units of TEM for comparisons. The purpose of this paper is to provide a method that allows measurement errors in large samples to be easily incorporated into the analyses.

This manuscript shows the effectiveness of using units of TEM in three different fields that use anthropometry. This was done to demonstrate the versatility of the method and strengthen the findings. During the review of the previous two manuscripts, a common question asked was ‘what are the error rates?’ At the time, it was sufficient to simply report the error rates. However, the authors wished to show the effectiveness of the method of ‘duplication’ by using units of TEM and comparing it to standard units (as used previously). This strengthened the ‘duplication’ method by showing that even when errors are taken into consideration, an individual can still be isolated from a population based on measurements. The method presented in this manuscript eliminates any questions regarding error rates associated with anthropometric measurement studies, more specifically, the previous two manuscripts.

This manuscript covers anthropometry as a whole, authors chose to use more than one example to show the versatility of the method, beyond forensic investigations.

Abstract

Anthropometry is a tool used in many fields. Anthropometrists make every effort to ensure that measurement errors are minimal, however, they cannot be eliminated entirely. Currently, measurement errors are simply provided alongside reported values. Measurement errors should be included into analyses of anthropometric data. This study proposes a method which incorporates measurement errors into reported values, replacing metric units with ‘units of technical error of measurement (TEM)’ for applications in forensics, industrial anthropometry and studies of biological variation.

The army anthropometry survey (ANSUR) contains 132 individual anthropometric dimensions of 3982 men and women. The concept of duplication and an Euclidean distance calculation were applied to the ANSUR for forensic–style identification analyses. A principal components analysis was applied to ANSUR as an example of a study of biological variation. The National Size and Shape Survey of Australia contains 65 anthropometric measurements of 1265 women. This sample was used to show how a woman’s body measurements could be ‘matched’ to clothing sizes from the garment industry. Euclidean distances show that the same person cannot be matched (>0) on measurements in millimetres but can in units of TEM (=0). Only 81 women can fit into any standard clothing size when matched using centimetres, with units of TEM, 1944 women fit. PCA shows little difference between using millimetres and units of TEM. The proposed method can be applied to all fields that use anthropometry. Units of TEM are proposed to be a more reliable unit of measurement for comparisons.

Key words: Physical anthropometry, Technical error of measurement, ANSUR

Introduction

Anthropometry refers to the measurement of the human body, in size, proportions and composition (Ulijaszek and Kerr 1999). Anthropometry is a standardised method which uses standardised instruments and techniques to take measurements of various parts of the body as accurately as possible. The information gathered from anthropometric surveys can be used across a wide range of fields including but not limited to: forensics, the garment industry, ergonomics, sports medicine, biological variation, health studies, and population studies.

Despite wide uses of anthropometry and the standardisation of the techniques, measurement errors still occur (Stomfai et al. 2011; Perini et al. 2005; Ulijaszek and Kerr 1999). Standardisation of the placement of anthropometric points, measuring equipment and of methods of taking measurements practically eliminated systematic errors. However, random overmeasuring or undermeasuring cannot be avoided since both the measurer and the subject are biological and thus variable in numerous ways. Acknowledging random errors is the only way to assure good quality data. Technical error of measurement (TEM) is the commonly used measure of precision within anthropometry (Mueller and Martorell 1988). The following formula can be used to calculate the TEM:

$$\mathrm{TEM} = \sqrt{\left(\sum \mathrm{D}^2\right)/2\mathrm{N}}$$

Where D is the difference between two measurements of the same dimension of the same individual taken on two occasions and N is the number of individuals so measured (Stomfai et al. 2011; Perini et al. 2005; Ulijaszek and Kerr 1999).

There are also other measures of error used such like: observer error, standard error of measurement (SEM), absolute difference and mean absolute difference (MAD) (Gordon et al. 2010).

The TEM has been deemed the most appropriate measure of error for this study because it assumes that errors occur from random events, as opposed to systematic error and bias, which other statistics such as the SEM measure. In large anthropometric surveys, much care is taken to reduce bias and systematic error as much as possible. All people taking measurements are highly trained in how to locate anthropometric points and measure between them. Standardised anthropometric equipment is also used which is designed to avoid any bias. Bias does not change the units of measurement, it changes the size.

Random variation of measurements of the same person may result from actual variation in body dimensions (eg diurnal variation in waist circumference), differences in measuring techniques used by different measurers and errors made by the same measurer on different occasions. TEM is a measure of variation of individual observations around actual values as they were at the moment of measurement. It may include intraobserver variability that is the difference in measurements taken of and by the same individual at different times and interobserver variability that refers to the difference in two or more measurements taken by more than one observer. TEM has an interpretation similar to that of a standard deviation. For example, if a measurement value was recorded as 140 mm with a TEM of 1.23 mm then the range of ±1.23 mm includes 68% of individuals whose actual measurement is 140 mm. Their recorded measurements will fall into the range from 138.77 mm to 141.23 mm. Due to its statistical nature, a TEM which has been calculated from a large study can be applied to an individual case.

Anthropometrists have concentrated their efforts on acknowledging measurement errors and reporting them alongside the ‘actual’ measurements (Perini et al. 2005) but they do not incorporate them into analyses of anthropometric data. This begs the question ‘Is simply acknowledging that an error exists and reporting it enough?’

This paper aims to introduce a new method which eliminates the ambiguity of analytical results based on measurements taken with random errors by replacing metric units of measurement with the units of TEM, a method which can be applied to all fields which use anthropometry. The advantages of using units of TEM will be explored by applying them to three examples of fields which use anthropometry: forensics, clothing industry and biological variation. The purpose of this paper is to provide a method that allows measurement errors in large samples to be easily incorporated into analyses. It is not the aim of this paper to report TEMs for various anthropometric surveys.

Materials

Two different anthropometric datasets were used to illustrate the versatility of the theoretical method proposed in this study. The first dataset used was an anthropometric study of U.S military personnel (ANSUR). ANSUR was used to show how TEMs can be used in craniometric comparisons and comparisons of human faces. The second dataset used was the National Size and Shape Survey of Australia, it was used to illustrate applications in the garment industry. Their descriptions are as follows:

ANSUR:

In 1988, 1774 men and 2208 women aged between 17 and 51 years were measured as part of a large anthropometric survey of U.S military personnel (ANSUR).

This study describes in detail the ancestry of all participants. However, considering only 5% of human variation occurs between populations (Cavalli-Sforza and Bodmer 1971), the sample was not divided by ancestry in each of the examples that used this database. The following head measurements were used: *Bitragion submandibular arc* – The surface distance between the right and left tragion across the submandibular landmark at the juncture of the jaw and the neck. *Bizygomatic breadth* – The maximum horizontal breadth of the face between the zygomatic arches. *Ear breadth* – The maximum breadth of the right ear perpendicular to its long axis. *Ear length* – The length of the right ear from its highest to lowest points on a line parallel to the long axis of the ear. *Ear protrusion* – The horizontal distance between the mastoid process and the outside edge of the right ear at its most lateral point (ear point). *Head breadth* – The maximum horizontal breadth of the head above the ears. *Head circumference* – the maximum circumference of the head above the supraorbital ridges and ears. *Head Length* – the distance from the glabella landmark between the brow ridges to opisthocranion. *Interpupillary distance* – the distance between the centres of the right and left pupils. These particular head measurements were chosen as the TEMs were reported. Although head/face measurements show less variability and errors when compared to measurements of the body (Lucas and Henneberg 2015a) the principle of this study remains the same. Authors chose to use head and face measurements to show the versatility of the method between different parts of the body and between fields. Head and face measurements were chosen to illustrate the application of this method to craniometric studies. Further details of the measurements of these distances are described the ANSUR manual (ANSUR, 1988).

TEMs were provided by a validation study applied to the ANSUR database (Gordon et al. 2010). The TEMs reported were interobserver, where 10 participants were measured twice in a day by different measurers. Although intraobserver variability is not reported separately, it is included in the interobserver measurement automatically as subjects are being measured multiple times. Therefore for use in the current paper, all TEMs will not be separated into inter and intra observer but will be regarded as one. TEMs were calculated separately for males and females and this was included in all analyses. The TEMs for males and females separately are reported in Table 1.

Table 1: The reported TEMs for both males and females used for the ANSUR database (ANSUR 1988, Gordon et al. 2010).

Dimension	TEM males	TEM females
Bitragion submandibular arc	2.27	2.34
Bizygomatic breadth	0.90	0.87
Ear breadth	0.80	1.22
Ear length	0.86	0.91
Ear protrusion	0.82	0.94
Head breadth	0.85	0.84
Head circumference	2.43	1.35
Head length	0.98	0.87
Interpupillary distance	0.50	0.68

National size and shape survey of Australia:

The National Size and Shape Survey of Australia was conducted in 2002. A total of 54 measurements were taken of 1265 women aged 18-70+years (Henneberg and Veitch 2003; 2005). Women were recruited during craft fairs in major capital cities around Australia and represent a wide range of variation. For the purposes of this paper, the following three measurements were used: *Bust circumference* - The horizontal circumference of the fullest part of the bust. *Hip circumference* – The maximum horizontal circumference of the lower trunk. *Waist circumference* - The minimum horizontal circumference of the trunk between the lower ribs and the iliac crest. The three measurements were chosen because they are commonly used in the garment industry to define “sizes” of female clothes. (eg bodices, blouses, jackets).

TEMs for this study were calculated from a set of measurements of 16 people, measured by 9 anthropometrists who took part in the survey. These were graduate students of human biology who before engaging in anthropometric surveys measured each other and some volunteers repeatedly. Since the National Size and Shape Survey of Australia was collecting data during craft fairs in capital cities across the country, settings of anthropometric stations were less rigorous than in the ANSUR survey, participants measured were allowed to wear an assortment of garments rather than standardised clothing, they were of widely differing ages and the range of variation of body shapes has not been limited in any way. The pre-survey re-measurements by members of the team were organised to mimic realistic conditions under which the survey was run. For consistency with the ANSUR database, the TEMs calculated were interobserver. The ANSUR team reported TEMs separately for each sex, because each ANSUR measurer was responsible for measuring only participants of the same sex ie. females measured females. This could have had some influence on the TEMs. The TEMs used for the National Size and Shape Survey of Australia were not separated by sex as the team who measured individuals consisted of both males and females and they measured individuals of both sexes since a small sample of males were also measured during the survey. The TEMs used in all analyses using the database of the National Size and Shape Survey of Australia are reported in table 2.

Table 2: The TEMs (mm) used for the National Size and Shape Survey of Australia.

Dimension	TEM
Bust circumference	46.75
Waist circumference	56.98
Hip circumference	26.95

Method

Units of TEM:

Instead of reporting the actual value of measurement in standard units with technical error specified as +/-, the actual dimension measured can be reported in units of TEM because any accuracy greater than TEM may be a random result not reflecting actual size. In order to convert the reported value, the 95% confidence range of this value is calculated as 1.96 times TEM, which in practice means dividing the measured value by two times TEM. Since the actual measurement cannot be made with precision greater than TEM, the result of the division, for simplicity's sake can be rounded to full integers of TEM. For example: If the reported value of ear length for an individual is 64 mm and the TEM of ear length is 0.86 mm for that study, then 64/(2*0.86)=37.209, it is then rounded to 37 units of TEM. This is sufficiently precise for analysing since TEMs are estimates based on relatively small samples.

We analyze anthropometric variables expressed in millimetres and in the units of TEM for differences they can produce in results of individual-to-individual comparisons using principal components analysis, Euclidean distances, and the exact matches on combinations of anthropometric diameters (search for duplicates, Lucas and Henneberg 2015a,b). We also apply TEM units for matching exact body dimensions of individuals with "garment size" dimensions specified as single values of chest, waist and hip circumferences on standard clothing size roll.

For individual-to-individual or individual-to a standard ("size") matches we use the following approach that produces a categorical answer: 'match' (value of 0) or 'no match' (value different from 0). The value '0' represents no actual difference between the two variables being measured, whether they be individual-individual or individual-to-a standard.

Lets characterize an individual *a* by a combination of *k* body dimensions and try to match this individual with another one *b* characterized by the set of *k* same dimensions d_1 , d_2 … d_k. We assume that the two individuals differ in their measurements of those dimensions as recorded in millimetres. We calculate a value *m* expressing the match of the dimensions of these two individuals where INT is the nearest integer.

$$m = \sum_{i=1}^{i=k} \mathrm{INT} \frac{|\mathrm{d_{ia}} - \mathrm{d_{ib}}|}{2\ \mathrm{TEM_i}}$$

IBM SPSS Statistics v 22 and Microsoft Excel are used for all statistical analyses.

Results

Example 1- Forensics:

In forensics, it is important to compare individuals for identification by matching their head or body dimensions. In such instances it is important to know how many characteristics are needed to confidently distinguish a given individual from all other individuals so that a wrong person is not identified. When comparing two images, or an image of a person with the living person, it is extremely important to know if they are a match and if so, are measurement errors the reason for the match. If an individual matches another while each measurement was taken with an error and they actually are not the same person, then this would be a devastating outcome. Not only for the accused but also for the person who conducted the analysis.

For the purposes of this example (Figure 1), males and females from the ANSUR database were mixed to make the sample larger. Sex is irrelevant in this case because distributions of male and female dimensions overlap and the goal of this analysis is to know if there are any two or more individuals who match one another on head/face measurements and if there is a difference in the number of matches between using millimetres and units of TEM. To answer this question, the concept of duplication as described in (Lucas and Henneberg 2015a) is used. Basically, a match is defined as any two measurements of the same characteristics eg. stature which are numerically identical. If two or more individuals match one another on all characteristics used for comparisons then they are considered duplicates. It is trivial to say that many people can be matched on head length only. When more measurements are added to a comparison, the probability of finding a matching individual decreases. To find how many measurements need to be used to find each individual to be different, characteristics will be added to the analysis in a stepwise approach until there are no duplicate matches and all individuals can be distinguished from each other. The list of head/face measurements (Table 3) was ordered largest to smallest in range as this produces the best results (Lucas and Henneberg 2015a).

Table 3: List of head/face dimensions from the ANSUR database and their range in mm and units of TEM for the sample of 3982 individuals.

Dimension	**Range (mm)**	**Range (units of TEM)**
Bitragion submandibular arc	138	32
Bizygomatic breadth	44	23
Ear breadth	24	21
Ear length	35	22
Ear protrusion	27	17
Head breadth	47	27
Head circumference	127	120
Head length	62	33
Interpupillary distance	26	40

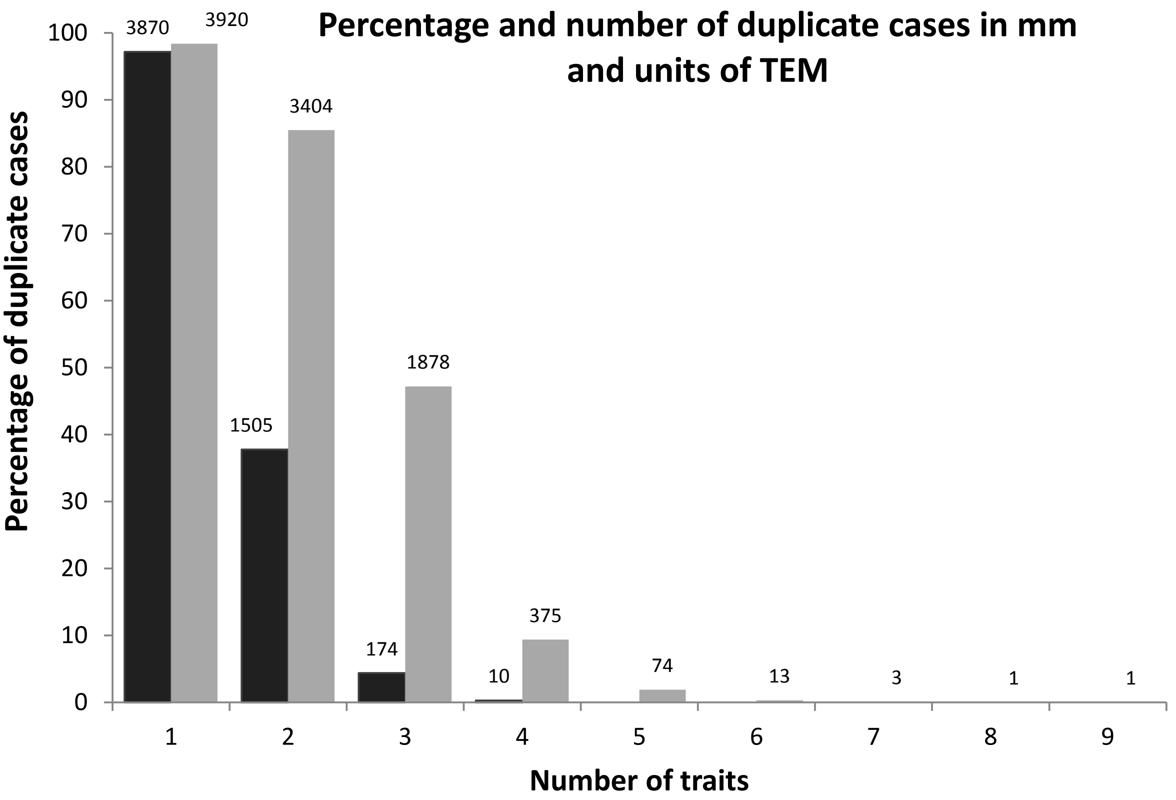


Figure 1: Percentage and number of duplicate cases of head/face measurements for males and females (N=3982) for millimetres and units of TEM.

Figure 1 shows the percentage and number of duplicate cases for males and females in both millimetres and units of TEM. In both millimetres and units of TEM, every time a characteristic is added, the number of duplicates decreases. However, it can be seen that the number of duplicates decreases slower when units of TEM are used, this is because the values are smaller and therefore there is a higher probability of having duplicates as there are less options for each measurement to fall into. It takes 5 measurements of the head/face to have no duplicates in the sample when millimetres are used. It takes more than 9 measurements of the head/face to find no duplicates in the sample when units of TEM are used. The case of the one remaining duplicate was investigated and it is two separate females, who are very similar on head/face measurements.

A Euclidean distance is commonly used calculation to compare two individuals to establish the level of similarity. The basic rule of the Euclidean distance is if two individuals are compared on any number of measurements and there are no differences between the values being compared then the distance will be zero, anything above zero indicates that two individuals compared are not the same person. The larger the value, the more differences there are between the two individuals. For the purposes of the example (Table 4), 20 males were chosen at random from the ANSUR database to illustrate the difference between using millimetres and units of TEM in an Euclidean distance matrix. Only one sex was chosen to show the similarities and differences, Males were chosen as they are more likely to commit crimes (Australian Bureau of Statistics 2013). In this instance we used a calculation of the average Euclidean distance that allows numerical comparisons of similarities among individuals who were characterised by somewhat varying number of dimensions since in practice it happens that a particular dimension of a particular individual cannot be measured (eg. on an image partly obscured by an object between the lens and the person). Thus the distance E is:

$$E = \frac{1}{k}\sum_{i=1}^{i=k}(d_{ia} - d_{ib})^2$$

Table 4: Euclidean distances between male subjects (N=22) for measurements of the head in both mm (bottom) and units of TEM (top).

Subject Number	17371	17710	9313	20899	11473	6194	10027	20969	18516	3575	16826	23804	2594	22929	10101	7739	1068	21749	17561	19995	21	22
17371	0	18	202	13	6	2	7	29	168	13	45	43	9	24	105	10	34	6	10	29	21	16
17710	143	0	33	18	17	19	7	7	24	14	113	6	45	17	289	18	5	23	11	8	66	63
9313	173	192	0	23	192	77	146	13	11	18	1617	8	720	31	2316	45	20	69	205	214	1313	1108
20899	49	258	238	0	9	13	17	23	37	16	220	23	91	8	489	8	27	18	9	30	127	111
11473	18	167	265	35	0	7	10	33	165	18	25	44	6	15	89	8	35	7	9	23	9	11
6194	22	104	98	75	56	0	4	13	69	5	143	18	28	19	272	13	16	7	9	33	92	68
10027	102	15	179	224	129	67	0	17	117	13	56	27	14	25	141	10	13	8	9	22	33	28
20969	163	37	94	288	223	98	41	0	8	8	416	3	153	15	722	18	6	5	16	30	298	249
18516	307	122	90	445	400	204	141	39	0	23	1325	9	595	32	1957	55	15	47	149	154	1057	899
3575	36	119	56	90	90	15	98	91	167	0	338	8	106	13	585	12	17	4	14	32	236	190
16826	66	192	399	113	32	111	146	283	497	173	0	561	3	335	3	52	414	151	51	61	4	7
23804	141	82	35	234	211	73	83	24	50	52	293	0	207	13	915	21	9	7	31	40	424	352
2594	35	124	309	94	21	72	93	198	389	113	17	215	0	108	9	20	156	18	4	25	6	8
22929	53	232	140	31	69	58	205	208	321	49	159	143	125	0	586	19	25	5	11	27	240	200
10101	100	160	449	198	71	153	122	263	476	214	26	314	24	240	0	185	723	260	105	120	4	8
7739	38	147	258	46	28	55	121	203	379	89	66	202	32	86	102	0	22	38	20	42	19	17
1068	259	52	158	404	313	169	56	27	50	180	365	71	282	332	331	297	0	11	24	30	302	252
21749	45	97	62	109	89	15	71	65	141	12	155	39	106	55	188	100	138	0	2	13	109	83
17561	31	79	141	91	50	20	55	83	202	36	82	78	45	61	104	52	163	19	0	17	36	31
19995	150	24	192	289	180	124	35	51	121	128	202	85	144	240	163	207	65	94	91	0	51	55
21	53	288	377	69	33	109	227	349	554	153	31	323	50	109	88	74	472	155	100	292	0	0
22	61	309	409	74	40	125	247	382	598	170	37	356	57	130	95	77	508	183	123	320	3	0

Note: The individuals who share the most similarity are boldfaced. The individuals who are matched with themselves are shaded. The individual who has been measured twice ('21' and '22') is outlined.

Table 4 shows the Euclidean distances between 20 male subjects from the ANSUR database chosen at random and a male individual who was measured twice. A male individual was measured once by each of the two authors and is included in the sample (‘21’ and ‘22’), making the total sample 22 males. The nine head measurements reported in the ANSUR sample were used except in the case of the male who was not from ANSUR, interpupillary distance was excluded as the measurement could not be accurately compared without a pupillometer. Head measurements were used as an example, as that is the part of the body that is often used in forensic cases of identification from images. The Euclidean distances between individuals have been calculated for millimetres (bottom of the matrix) and units of TEM (top of the matrix). The male who was measured twice is number 21 and 22. There is a distance of 3 between this individual and himself when measured in millimetres. This is clearly different from zero so that 21 should be considered as a different person from 22, at least initially. Further investigation will be required to establish their identity. When measured in units of TEM the difference is zero. A clear answer. In all other cases, no two males have a zero result to be considered the same. This should be so because we chose different individuals from the ANSUR database and our previous analyses indicated that when 8 or more measurements of the head and face are used in millimetres or units of TEM, there are no duplicate males (Figure 1).

Example 2 - Clothing industry:

A mentioned above, it is the goal of the clothing industry to design products which fit a large number of people. Clothing comes in standard sizes which are aimed to fit most people satisfactorily. With online shopping being a new concept in today’s society, many women do not get the opportunity to try on clothing. Clothing is often ordered from overseas where the sizing standards are different. Each woman simply wants to know, ‘will I fit into this garment or not?’ and ‘how much of a difference am I willing to accept?’. To answer these questions the ‘match’ or ‘no match’ categories were calculated for each of the 1265 women in the National Size and Shape Survey of Australia. This dataset was used as it was originally designed for the clothing industry and thus had the measurements needed to use clothing as an example. It was also used to construct a sizing system following the old common practice of dividing the sample into groups based on one measurement, in this case the bust circumference, categorised into groups differing by a set amount and then calculating averages of other dimensions for each group. In this case we used 50 mm steps for bust circumference. Examples from the standard size roll are presented in Table 5.

Table 5: Examples from the standard size roll used in matching analyses for bust, waist and hip circumference.

Standard size	Bust circumference (mm)	Waist circumference (mm)	Hip circumference (mm)
Size 10	800	670	920
Size 16	950	820	1050
Size 18	1000	870	1090
Size 26	1200	1080	1260

Average dimensions of the women in the National Size and Shape Survey of Australia are published in (Henneberg and Veitch 2003, 2005, Henneberg and Ulijaszek 2010).

Table 6: The number of women whose body measurements match each standard sized garment when measured in centimetres and units of TEM.

Standard size	**Number of women who match (cm)**	**Number of women who match (units of TEM)**
10	2	216
12	9	317
14	20	354
16	21	343
18	9	266
20	9	206
22	9	134
24	1	71
26	1	37
Total	81	1944

Table 6 shows the number of women whose body dimensions match each Australian standard clothing size when the measurements are taken in centimetres and units of TEM. It can be seen that very few women match the standard clothing size in centimetres, though we allowed for a one centimetre read-out error. However, more than the number of women measured (N=1265) matched the standard clothing sizes in units of TEM. This shows that some women may fit into more than one standard size of clothing if errors are accounted for (Table 6).

Table 7: An example of a woman who does not match a standard clothing size in millimetres but does match in units of TEM.

Clothing size	**10**	**12**	**14**	**16**	**18**	**20**	**22**	**24**	**26**
Difference in mm	14	7	1	4	12	18	24	31	40
Difference in units of TEM	1	**0**	**0**	2	3	5	7	9	13

Note: matches as represented by the value ‘0’ are boldfaced. Other values are a sum of differences from the standard.

Table 7 is an example of a particular woman who does not match a standard clothing size in millimetres. The same woman matches a number of different standard clothing sizes (12 and 14) when measured in units of TEM. The further away from a 'match' the woman is, the larger the difference (value). In this example, she can fit into a size 12 and 14 but if one other standard size had to be chosen, she would fit better in a size 10 than she would into a size 16. This is because the difference between her body measurements (when errors are accepted) and the standard size 10 is less than the size 16.

Example 3- Biological variation:

In biological anthropology, a principal components analysis is used to illustrate similarities and differences between skulls of different time periods based on their craniometric characters. An attempt is sometimes made at discerning groups of skulls plotting close together to assign them to a taxonomic group (eg Bookstein et al. 1999;White et al. 2003). Such attempts often do not produce clear results because human variation is continuous and substantial overlaps between skulls from different populations occur. Although the ANSUR database only contains measurements from living modern humans, the nine head dimensions mentioned above can still be used as an example to show similarities or differences between individuals characterised by craniometric dimensions recorded in millimetres and units of TEM. In this case only the males were used as an example, to avoid any differences that are caused by sexual dimorphism.

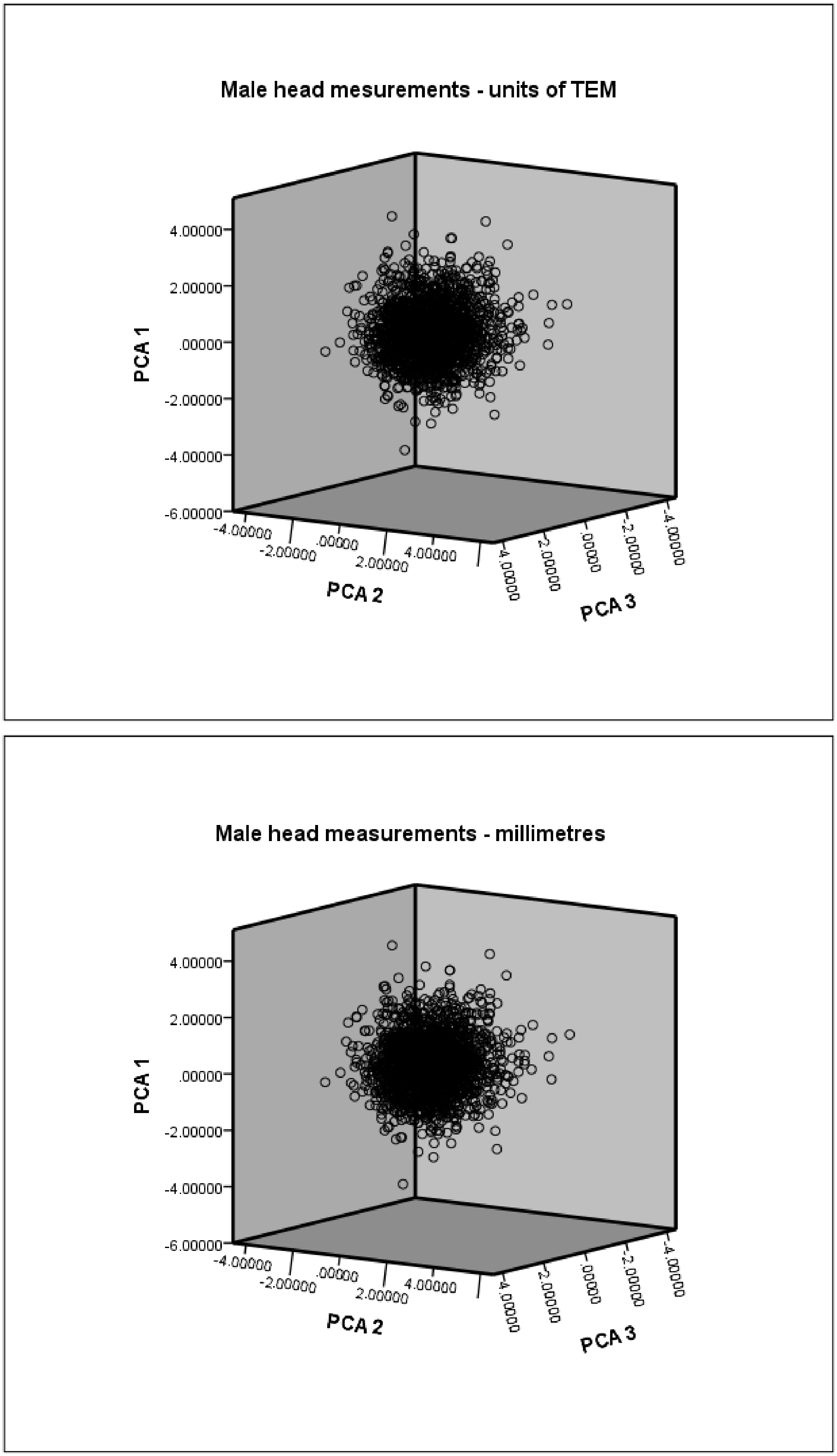


Figure 2: Principal components analysis for male head measurements in millimetres and units of TEM.

Figure 2 is an example of a principal components analysis conducted on measurements of the head. As mentioned previously, there was no separation of the data into ancestral groups. As could be expected, these results show no separate groups of male modern humans (Cavalli-Sforza and Bodmer1971) ie. there is no separation of groups which can be attributed specific ancestry. These results show continuous variation in both millimetres and units of TEM. For the purposes of this paper it is irrelevant whether some separation into ancestry groups occurred or not. What is important is the fact that millimetres and TEMs give the same result. Lack of clear subdivision into groups is a biological fact of human variation (Cavalli-Sforza and Bodmer 1971). The fact that there are no substantial differences shows that using units of TEM instead of millimetres will not compromise the data by showing false similarities or differences between individuals.

Discussion

In the case of the duplicate females from the ANSUR database, it was found that these are in fact two separate females. The reason that they are considered ‘duplicates’ of one another is that they have very similar measurements, however, not identical in millimetres, only in units of TEM. The reason why they are the same in units of TEM is because the TEM has a range of millimetres. In a practical situation, in order to distinguish between the two women adding more measurements should prove successful. If this were a case where only selected measurements could be taken from an image due to distortion or other factors which influence visibility of characters then using descriptives as well as metrics to distinguish between the two women would be practical. Finding two females who match one another in units of TEM illustrates the proposed method perfectly. Simply finding less matches when the measurements are expressed in millimetres is not sufficient in a practical setting as errors are not considered. Two sets of measurements of the same person may differ by some millimetres due to random errors. Considering this difference as real could lead to devastating consequences. Using units of TEM eliminates any doubt that may be caused by simply reporting the errors alongside the reported measurements.

This method has particular practical use for automatic search through large samples for duplicates. If comparing two or three individuals, simply reporting the range of errors may be sufficient as it can be established whether or not the measurements are within error ranges by simply observing the values. However in large surveys, this is not practical.

The principal components analysis (Figure 2) showed no substantial differences between using millimetres and units of TEM. By simply converting millimetres to units of TEM all of the dimensions of individuals are comparable in a similar way though their numerical values differ. The small differences between millimetres and units of TEM that can be seen are due to the difference in numbers of units in each dimension that depends on both, the size of the dimension and its TEM.

The level of accepted error may change for each application. In other words, the calculation does not always have to use 95% confidence range. It may include larger error ranges. This would be most applicable to the ergonomics industry, which like the clothing industry, designs furniture or workspaces to fit a great number of individuals. However, the difference lies in the number of products for the same number of individuals. For example there may only be one design of a particular chair, this chair would fit the same number of people that nine different standard clothing sizes would fit because it is not designed to fit perfectly and it may be adjustable.

The proposed method is a general method which can be applied to all fields which use statistical analysis of any measurements taken with an error. The actual value of a measurement is not the number of standard SI units, but the number of TEM units included in the measurement. This paper shows some possible uses and advantages of using units of TEM as opposed to using millimetres and simply reporting TEMs.

Acknowledgements

The authors would like to thank Daisy Veitch, the Managing Director of the SHARP Dummies Pty. Ltd. that co-funded and co-directed the National Size and Shape Survey of Australia, without whose input in data collection this publication would not be possible.

References

ANSUR database, 1988, Clauser, CE, Tebbetts IO, Bradtmiller, B, McConville, JT, Gordon, CC Measurer's Handbook: US Army Anthropometric Survey 1987-1988, Technical Report NATICK/TR-88/043. Gordon, CC, Bradtmiller, B, Churchill, T, Clauser, CE, McConville, JT, Tebbetts, IO, Walker, RA *1988.* Anthropometric Survey of US Army Personnel: Methods and Summary Statistics, Technical Report NATICK/TR-89/044, United States Army Natick Research, Development and Engineering Center, Natick, Massachusetts.

Australian Bureau of Statistics 2013, ‘Gender indicators, Australia’, cat. no. 4125.0, ABS, Canberra.

Berry, KM, Hennes, L 2008, *The Fashion Design System; Pattern engineering for the 21st century,* The Fashion Design System, Nambour.

Bertillon, A 1886, *Identification anthropometriques. Methode nouvelle de determination de L'iidentite individuelle*, Masson, Paris.

Bertillion, A 1890, *La Photographie judiciaure, avec un, appendice sur la classification et l'identification anthropometrique*, Gauthier-Villars, Paris.

Bookstein, F, Schäfer, K, Prossinger, H, Seidler, H, Fieder, M, Stringer, C, Weber, GW, Arsuaga, JL, Slice, DE, Rohlf, FJ, Recheis, W, Mariam, AJ, Marcus, LF 1999, 'Comparing frontal cranial profiles in archaic and modern HOMO by morphometric analysis', *Anatomical record*, vol. 257, pp. 217-224.

Cavalli-Sforza, LL, Bodmer, WF 1971, *The genetics of human population*. WH Freeman and company, San Francisco.

Das, B, Sengupta, AK, 1996, 'Industrial workstation design: A systematic ergonomics approach', *Applied Ergonomics*, vol. 27, pp. 157-163.

Gordon, C, Claire, C, Blackwel,l CL, Bradtmiller, B, Parham, JL, Hotzman, J, Paquette, SP, Corner, BD, Hodge, BM 2010, *Anthropometric Survey of US Marine Corps Personnel: Methods and Summary Statistics*. No. NATICK/TR-13/018. Army Natick Soldier Research Development and Engineering Center MA, 2013.

Henneberg, M, Miguel, CD 2004, 'Hominins are a single lineage: brain and body size variability does not reflect postulated taxonomic diversity of hominins', *HOMO*, vol. 55, pp. 21-37.

Henneberg, M, Veitch, D 2003, 'National size and shape survey of Australia', Kinanthreport. 16:34–39.

Henneberg, M, Veitch, D 2005, ‘Is obesity as measured by body mass index, and waist circumference in adult Australian women 2002 just a result of the lifestyle?’, *Journal of Human Ecology (Special Issue),* vol. 13, pp. 85–89.

Henneberg, M, Ulijaszek, S 2010, ‘Body frame dimensions are related to obesity and fatness: lean trunk size, skinfolds and body mass index’, *American Journal of Human Biology*, vol. 22, pp. 83-91.

Howells, WW 1973, ‘Cranial variation in man: a study by multivariate analysis of patterns of difference among recent human populations’, Papers of the Peabody Museum no.67. Cambridge MA: Peabody museum, Harvard University.

Howells, WW 1989, ‘Skull shapes and the map: craniometric analyses in the dispersion of modern HOMO’, Papers of the Peabody Museum no.67. Cambridge MA: Peabody museum, Harvard University.

Lucas T, Henneberg, M 2015a, ‘Comparing the face to the body, which is better for identification?’, *International Journal of Legal Medicine*. DOI: 10.1007/s00414-015-1158-6.

Mueller, WH, Martorell, R 1988, *Reliability and accuracy of measurement. In Anthropometric Standardisation Reference Manual*, Champaign, IL: Human Kinetics Books.

Perini, TA, Oliveira, GLD, Ornellas, JDS, Oliveira, FPD 2005, ‘Technical error of measurement in anthropometry’, *Revista Brasileira De Medicina Do Esporte*, vol. 11, pp. 86-90.

Salvendy, G 2012, *Handbook of human factors and ergonomics*. John Wiley & Sons.

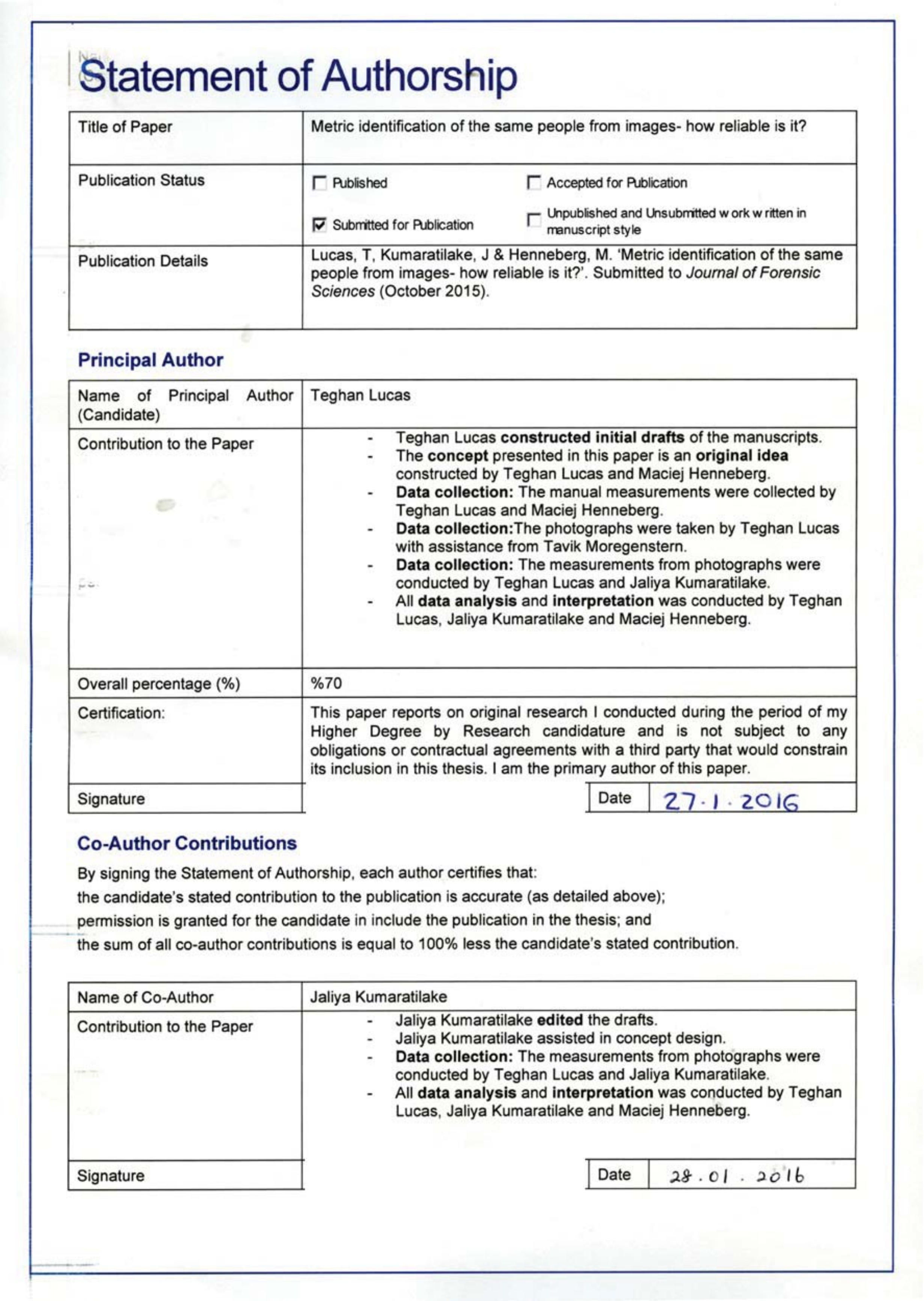
Scoleri T, Lucas T, Henneberg, M 2014, ‘Effect of garments on photoanthropometry of body parts: Application of stature estimation’, *Forensic Science International*, vol. 237, pp.1-12.

Stomfai, S, Ahrens, W, Bammann, K, Kovács, E, Mårild, S, Michels, N, Moreno, LA, Pohlabeln, H, Siani, A, Tornaritis, M, Veidebaums, T, Molnár, D 2011, ‘Intra- and inter-observer reliability in anthropometric measurements in children’, *International Journal of Obesity*, vol. 35, pp. 545-551.

Ulijaszek, SJ, Kerr, DA 1999, ‘Anthropometric measurement error and the assessment of nutritional status’, *British Journal of Nutrition*, vol. 82, pp. 165-177.

White, TD, Asfaw, B, DeGusta, D, Gilbert, H, Richards, GD, Suwa, G, Howell, FC 2003, ‘Pleistocene HOMO Sapiens from Middle Awash, Ethiopia’, *Nature*, vol. 423, pp.742-747.

Statement of Authorship Manuscript 5:



Statement of Authorship

Title of Paper	Metric identification of the same people from images- how reliable is it?
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Lucas, T, Kumaratilake, J & Henneberg, M. 'Metric identification of the same people from images- how reliable is it?'. Submitted to *Journal of Forensic Sciences* (October 2015).

Principal Author

Name of Principal Author (Candidate)	Teghan Lucas
Contribution to the Paper	- Teghan Lucas **constructed initial drafts** of the manuscripts. - The **concept** presented in this paper is an **original idea** constructed by Teghan Lucas and Maciej Henneberg. - **Data collection:** The manual measurements were collected by Teghan Lucas and Maciej Henneberg. - **Data collection:**The photographs were taken by Teghan Lucas with assistance from Tavik Moregenstern. - **Data collection:** The measurements from photographs were conducted by Teghan Lucas and Jaliya Kumaratilake. - All **data analysis** and **interpretation** was conducted by Teghan Lucas, Jaliya Kumaratilake and Maciej Henneberg.
Overall percentage (%)	%70
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.

Signature		Date	27.1.2016

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

the candidate's stated contribution to the publication is accurate (as detailed above);

permission is granted for the candidate in include the publication in the thesis; and

the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Jaliya Kumaratilake
Contribution to the Paper	- Jaliya Kumaratilake **edited** the drafts. - Jaliya Kumaratilake assisted in concept design. - **Data collection:** The measurements from photographs were conducted by Teghan Lucas and Jaliya Kumaratilake. - All **data analysis** and **interpretation** was conducted by Teghan Lucas, Jaliya Kumaratilake and Maciej Henneberg.

Signature		Date	28.01.2016

Name of Co-Author	Maciej Henneberg
Contribution to the Paper	- Maciej Henneberg **edited** the drafts. - The **concept** presented in this paper is an **original idea** constructed by Teghan Lucas and Maciej Henneberg. - **Data collection:** The manual measurements were collected by Teghan Lucas and Maciej Henneberg. - All **data analysis** and **interpretation** was conducted by Teghan Lucas, Jaliya Kumaratilake and Maciej Henneberg.
Signature	Date 27.1.16

Please cut and paste additional co-author panels here as required.

5. Metric identification of the same people from images, how reliable is it?

Teghan Lucas[a*], Jaliya Kumaratilake[a] and Maciej Henneberg[a]

[a] Biological and Comparative Anatomy Research Unit, Medical School, University of Adelaide, Adelaide 5005, Australia.

Submitted to Journal of Anthropology 25th April 2016.

[*]Corresponding author.

Teghan Lucas

Medical School North Building, Frome Road, The University of Adelaide, South Australia 5005, Phone: +61 8 8313 6324. Email address: teghan.lucas@adelaide.edu.au

Context

Much of the research that investigated identification from images was conducted by experts in computer analysis (BenAbdelaker and Davis 2006; BenAbdelaker and Yacoob 2008; Scoleri and Henneberg 2012; Scoleri, Lucas and Henneberg 2014), not biological anthropologists or those with an extensive knowledge of the human body. Therefore, the effectiveness of the use of facial and body measurements as a biometric tool has not been explored in reference to identification from images. This was addressed by the findings described in manuscripts 2 and 3. We have shown that knowledge of anatomy and anthropometry reduces errors associated with taking measurements from images (manuscript 1).

The techniques proposed by computer analysis experts include complex formulae which are tedious and time consuming to the operator. These techniques have an error rate of 5% (Scoleri, Lucas and Henneberg 2014), which is unacceptable for identifications in a court of law.

The aims of this research were to use a mathematical approach known as the ‘anharmonic ratio’, which eliminates effects of image distortions to isolate individuals from the sample (n = 20). This method was chosen as it has been previously applied to images of inanimate objects with success; it is quick and does not require computer analytical techniques. Therefore, persons with extensive knowledge of the human body (who may not necessarily have computer analysis knowledge) can use it and eliminate any errors caused by lack of anatomical knowledge of computer operators that have been outlined in manuscript 1.

Results showed that proportions derived from use of the anharmonic ratio were not sufficient to correctly identify individuals as error rates multiply with the use of body ratios.

In previous literature, the anharmonic ratio was successful in identifying objects. However, previous studies, which used ratios of the face could not identify the individuals, this was consistent with these findings. This paper was built on previous studies by using ratios of the body, as well as the face measurements as described in manuscript 3, which outlined the usefulness of body measurements for the identification of individuals. However, results show that there is little difference between the use of the face and body measurements when ratios are applied.

The findings of this study outline the need for further multidisciplinary research between image/computer analysis experts and biological anthropologists.

Abstract

Ratios have been applied to humans to identify individuals from images. These attempts have been proven unsuccessful, as camera angle, height and distortions of the image affected the results. The anharmonic ratio is a ratio of ratios, it has proved successful in the identification of objects from images, as it is not affected by any distortions. The anharmonic ratio was applied to the human body and face to identify individuals from their images. Twenty South Australian males aged 16 – 65 years faces and bodies were measured using standard anthropometric techniques. Participants were photographed in high quality images and recorded by standard surveillance camera (low quality images). Ten ratios were calculated from manual measurements and from all images. An Euclidean distance showed ratios incorrectly identified individuals 64.3% of the time between images of different quality. Variation of ratios between individuals is low so that standard deviations of ratios are of the magnitude similar to technical errors of measurements, Therefore participants cannot be isolated based on ratios. Ratios are an unreliable method for identification.

Keywords: Forensic science, anharmonic ratio, TEM, identification, anthropometry, images.

Introduction

Criminal activities are increasingly being recorded in closed circuit television systems (CCTV). Scientists with expertise in biological anthropology or computer image analysis are often called upon to identify the individuals from the images. Image identification involves comparing anatomical features seen on the image, of the person who committed the crime (person of interest) to the person who is suspected of committing the crime (suspect). Often, the images that are available for the analysis are of different quality (i.e high or low, particularly in relation to clarity), as they are commonly taken at different times and by different cameras It is not uncommon to have both high and low quality images in one case.

In order to accept a testimony of identification in a court of law, it should adhere to the Daubert Criteria (Ireland and Beaumont 2015; Crumbley and Cheng 2014; Grivas and Komar 2008). The Daubert criteria aim to make testimonies of expert witnesses as reliable as possible by ensuring that they are peer reviewed, repeatedly tested, have a known error rate and are based on sufficient facts or data.

As a result of the Daubert Criteria, the interval scale of measurement is considered to be superior to the categorical scales (Edmond et al. 2009), which are currently used by expert biological anthropologists. Interval scales of measurements could be expressed in any units that are directly measured, for example, nose width is 68 mm if a millimetre is used as the interval. Categorical scales of measurement, use adjectives to describe anatomical variations, for example, the nose is wide.

In the past, the evidence/opinion presented by an expert witness for the identification of an individual from images by using categorical classifications has been criticised as unreliable (Edmond 2008; Edmond et al. 2009) and the assessment of images in categorical scales is open to interpretation (Biber 2009). Therefore, research has been focused on the application of interval scale measurements to images for the identification of individuals (BenAbdelaker and Davis 2006; BenAbdelaker and Yacoob 2008; Scoleri and Henneberg 2012; Scoleri, Lucas and Henneberg 2014). It has the advantage of reproducibility and the ability to calculate error rates (Scoleri et al. 2014). Disadvantages of the methods that use interval scale measurements are that they require complex analytical formulae (BenAbdelaker and Davis 2006; BenAbdelaker and Yacoob 2008; Scoleri and Henneberg 2012; Scoleri, Lucas and Henneberg 2014) and the processes are tedious, thus time consuming. Taking reliable interval scale measurements from 2D images is difficult because of perspective (angular distortion) and possible rectilinear and curvilinear distortions that could result from the quality of the video equipment's used (Chen et al. 2010). Another source of error is introduced from the pixelation of digital images, which particularly results from the enlargement of small images (eg. enlargement of the image of a face to obtain the details). The lowest error rates that are scientifically accepted with these methods lie at 5% (Scoleri et al. 2014). However, 5% of stature is approximately 80mm, which is unacceptable in a court of law as it broadens the suspect pool significantly. Therefore, an error rate of 5% does not allow accurate comparisons between images.

To avoid complexities of correcting for angular and rectilinear distortions, various authors have used facial proportions measured on photographs taken in standard position in an attempt to identify known individuals from images (Kleinberg and Siebert 2012; Kleinberg, Vanezis and Burton 2007; Moreton and Morley 2011). Kleinberg and Siebert (2012) investigated the use of ratios in facial identifications. It was found that their techniques were not able to sufficiently isolate an individual from the sample. Kelinberg, Vanezis and Burton (2007) reported similar results, showing that individuals could only be correctly identified approximately 22%-25% of the time at best, even under the most optimal conditions. In each of these cases the images were of a high quality, which does not represent real life forensic conditions. Moreton and Morley (2011) studied the effects of image quality as well as angle of the camera, distance and lighting. They found that proportionality indices changed significantly with camera angles, distance, resolution and the lighting conditions. The conclusions of these studies were similar (Kleinberg and Siebert 2012; Kleinberg, Vanezis and Burton 2007; Moreton and Morley 2011). and indicated that proportions of facial measurements are not adequate for the identification of an individual from a sample. However, the proportions used were single proportions eg. taken as a percentage ratio of one trait to another, therefore these proportions are often subject to effects caused by image conditions, as reported. The anharmonic ratio is an improvement of the traditional proportional measurement used in previous investigations (Kleinberg and Siebert 2012; Kleinberg, Vanezis and Burton 2007; Moreton and Morley 2011). The anharmonic ratio is a ratio of ratios, thus it is not affected by any effects that camera angle, and some other photographic distortions may have on the identification process

The current paper aims to isolate individuals using the anharmonic (cross) ratios of the body and the face from images of different quality. The body and the face will be used as both are extremely variable and can be used for the identification of an individual Lucas and Henneberg (2015a). For the anharmonic ratio to be successful as an identification technique, it must isolate an individual from the sample, i.e. should not allow to find any duplicates (Lucas and Henneberg 2015a,b) based on the ratios. A secondary aim of the paper is to provide error rates for the measurements taken between anthropometric points placed on images of different quality.

Materials and methods

Principles of the method:

The anharmonic (cross) ratio is a ratio of ratios. It has been previously applied to photographs for identification purposes (Fryer 2000). The ratio remains a constant for a particular object through projective projection i.e. remains the same irrespective of the angle and the height of the camera or the size of a photograph. Therefore, the anharmonic ratio is ideal for comparisons of proportions between images. Although it is an ideal method for the identification of objects from photographs, and adheres well with the Daubert criteria, anharmonic ratio has not been applied to the human body. The method relies on identifying four collinear points (A,B,C,D). The anharmonic ratio of distances between those four points, is defined as:

$$\Re\ (A,B;C,D) = \frac{AC}{BC} \Big/ \frac{AD}{BD}$$

An example is shown in Figure 1. Points of nasion, subnasale, stomion and gnathion are collinear as are vertex, nasion, subnasale and stomion. The example illustrates how the same points can be used in different ratios as long as they remain collinear.

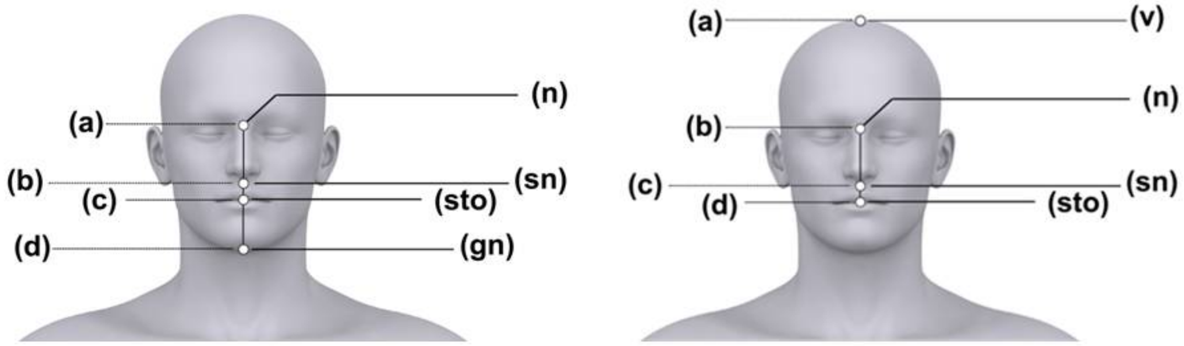


Figure 1: Example of four collinear points of the face and head. The example illustrates how the same points can be used for different ratios. The collinear points of the face and head, vertex (v) nasion (n), subnasale (sn), stomion (sto) and gnathion (gn).

If points are not already collinear, they can be realigned to become collinear with other points. A point may be moved horizontally or vertically in the same plane to align with other points, as long as the distance between the two points does not change due to the realignment. Figure 2 shows the upper body of an individual with four points marked along the horizontal line marked a, the two points marked ‘z and y’ can be realigned with the two points marked along the line ‘a’ by moving them straight down.

When applying these principles to the human body, only points whose anatomical position is fixed can be used. Distances should not be measured between points whose anatomical relationship can be altered by altering position of the body.

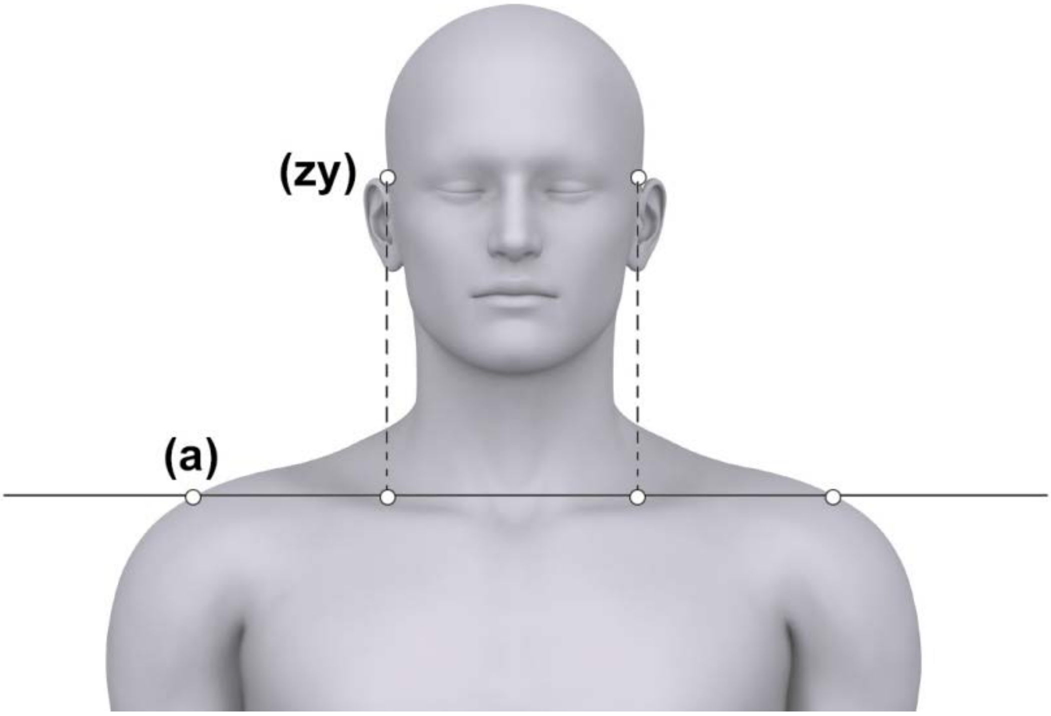


Figure 2: realignment of pre-existing points along a vertical axis to become collinear with points along a horizontal line.

The number of ratios that can be applied to the human body can exceed the number of fixed anatomical points. The reason for this being, that a particular point may be used in more than one ratio. In two different ratios , a fixed anthropometric point such as symphysion is used as either B or C (Table 1) If a point is labelled as A for one ratio, it does not necessarily have to be point A in another (eg nasion in ratios 2 and 3 Table 1).

Application of the method:

Twenty males between the ages of 16 and 65 years were recruited from South Australia. Like in the other studies (Kleinberg and Siebert 2012; Kleinberg, Vanezis and Burton 2007; Moreton and Morley 2011) males were chosen as they are the most likely sex to commit crimes that are caught on video surveillance systems (ABS 2013).

Three experience anatomists were asked to locate anthropometric points on the participants bodies and measure the distances between them. A set of standardised anthropometric measurements as defined by Martin and Saller (1957) were manually taken from each of the males using GPM anthropometer, sliding and spreading calipers (Table 1). Males were measured wearing shorts only. All measurements/landmarks were chosen as they are visible anteriorly (Figure 3).

Table 1: Anthropometric measurements/landmarks and their associated definitions taken manually of each of the male participants.

Measurement	Landmarks	Definition
Stature	b-v	The distance between the floor (b) and the highest point on the top of the skull when the head is held in Frankfurt horizontal (v).
Tragion height	b-t	The distance between the floor (b) and the superior point on the tragus of the ear (t).
Acromiale height	b-a	The distance between the floor (b) and the acromiale (a), the point at the superior and lateral borders of the acromion process.
Suprasternale height	b-sst	The distance between the floor (b) and the suprasternale (sst), the lowest point in the suprasternal notch in the midsagittal plane
Iliocristale height	b-ic	The distance between the floor (b) and the iliocristale (ic), the highest and the most lateral palpable point of the iliac crest of the pelvis
Trochanterion height	b-tro	The distance between the floor (b) and the trochanterion (tro), the superior point of the greater trochanter of the femur
Symphyseal height	b-sy	The distance between the floor (b) and the symphysion (sy), the point on the superior margin of the pubic symphysis in the midsagittal plane
Tibiale height	b-ti	The distance between the floor (b) and the tibiale laterale (ti), the superior point on the lateral condyle of the tibia.
Bi-acromial width	a-a	The distance between the left and right acromiale landmarks.
Bi-iliocristal width	ic-ic	The distance between the right and left iliocristale landmarks.
Bizygomatic width	zy-zy	The maximum horizontal breadth between the zygomatic arches.
Metopion- nasion length	m-n	The distance between the metopion (m), the intersection between median sagittal plane and horizontal line between left and right frontal eminences and nasion (n) the point where the midsagittal plane crosses the junction between the frontal and nasal bones, the deepest root of the nose.
Metopion-subnasale length	m-sn	The distance between the metopion (m), the intersection between median sagittal plane and horizontal line between left and right frontal eminences and subnasale (sn) the point where the nasal septum meets the philtrum.
Metopion-stomion length	m-sto	The distance between the metopion (m), the intersection between median sagittal plane and horizontal line between left and right frontal eminences and stomion (sto) the midpoint of the occlusal line between the lips.
Metopion-gnathion length	m-gn	The distance between the metopion (m), the intersection between median sagittal plane and horizontal line between left and right frontal eminences and gnathion (gn) the most inferior point on the body of the mandible in the midsagittal plane

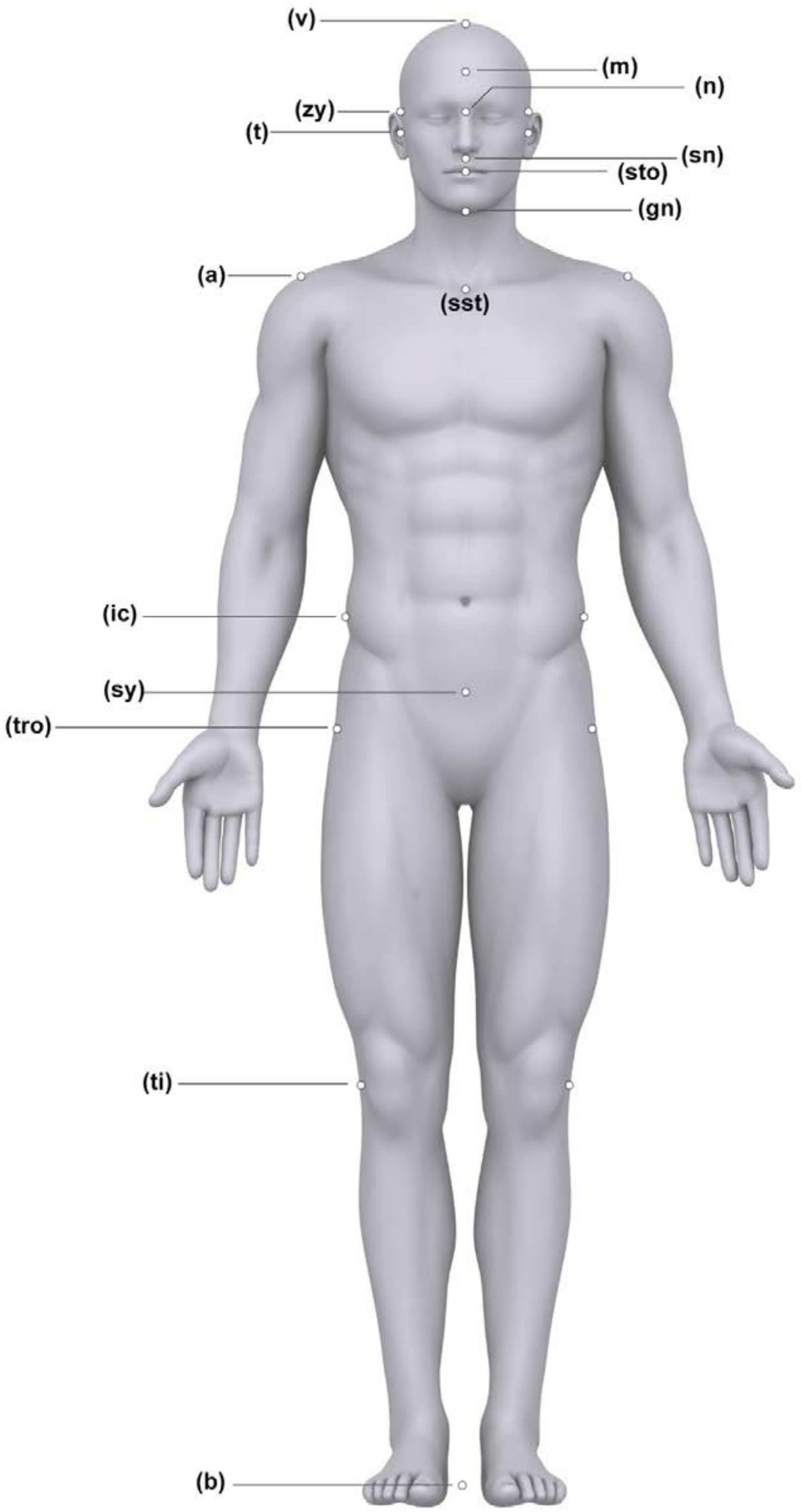


Figure 3: Anterior view of the entire body showing: (v) vertex, (m) metopion, (n) nasion, (zy) zygion (left and right), (t) tragion (left and right), (sn) subnasale, (sto) stomion, (gn) gnathion, (a) acromiale (left and right), (sst) suprasternale, (ic) iliocristale (left and right), (sy) symphysion, (tro) trochanterion (left and right), (ti) tibiale laterale, (b) base. Please note the point for zygion is slightly misplaced due to the point for tragion.

Participants were then photographed using a Panasonic Lumix DMC camera in the anatomical position wearing shorts only. The distance between the camera and the participant was 6m. The camera has a resolution of 12 megapixels with a 1.5x crop factor compared to a 35mm camera. The lens was a 14mm-45mm 3.5 zoom lens which is equivalent to a 28-90mm lens on a standard 35mm camera. All images were taken at 35mm, equivalent to 70mm on a standard 35mm camera. These photographs of participants are considered to be of a high quality as all details of the face and body are easily visible, while distortions are minimised. Therefore all photographs of this quality are referred to as ‘high quality’ in this paper.

Participants were then videoed using an Axis 216MFD (CCTV) at the University of Adelaide Medical School, foyer corridor of 8 m in length, 2 m in width and 3 m in height. The camera records video with a resolution of 1280x1024 pixels at a rate of 5 frames per second. The focal length is between 2.8mm and4.0mm. The mounting height of the camera is between 2.7m and 3.0metres off the ground. This is consistent with standard CCTV surveillance systems. Participants were pictured wearing a pair of trousers and a shirt. Participants wore every day clothing to represent real life forensic scenarios. Clothing is shown to have little effect on assessment of body shape (Lucas, Kumaratilake and Henneberg 2014). Participants were asked to walk the length of the corridor and pause on a marker line placed on the floor about 5 m away and facing towards the camera. This position is comparable with the position the participants were in, in the high quality photographs. These photographs of participants are considered to be of a low quality as details of the face and body are not clearly visible. Therefore, all photographs of this quality are referred to as ‘low quality’ in this paper. For a comparison of high and low quality images used in analysis refer to figure 4.

Figure 4: A comparison of high and low quality images. Note: the participants face has been blocked out for confidentiality reasons.

All previously defined landmarks were marked on both the high and low quality images of the participants. The images were printed on photographic paper and landmarks were marked manually together with distances between landmarks. A total of 10 anharmonic ratios were calculated for the body and face of each participant using the landmarks (Table 2).

Table 2. The ratios and corresponding landmarks applied to each of the participants (n=20).

	A	**B**	**C**	**D**
Ratio 1	Acromiale	Trochanterion	Tibiale	Base
Ratio 2	Metopion	Nasion	Subnasale	Stomion
Ratio 3	Nasion	Subnasale	Stomion	Gnation
Ratio 4	Acromiale	Symphysion	Tibiale	Base
Ratio 5	Vertex	Suprasternale	Symphysion	Base
Ratio 6	Tragion	Suprasternale	Symphysion	Base
Ratio 7	Vertex	Tragion	Suprasternale	Base
Ratio 8	Acromiale(L)	Iliocristale(L)	Iliocristale (R)	Acromiale (R)
Ratio 9	Iliocristale (L)	Zygomatic (L)	Zygomatic (R)	Iliocristale (R)
Ratio 10	Acromiale (L)	Zygomatic (L)	Zygomatic (R)	Acromiale (R)

In order to apply the same ratios used in photographic analyses to the manual measurements, the distances between landmarks were calculated by subtracting one measurement from another, i.e. the distance AC for ratio 1 (table 2) is the distance between acromiale and tibiale. This was not measured directly. The distance, tibiale to base was subtracted from the distance acromiale, in order to obtain the distance acromiale to tibiale. This allowed an accurate comparison to be made between each of the three settings; manual measurements, and the measurements from the high and the low quality photographs.

In all three situations, measurements were repeated separately by another anthropometrist. This allowed calculations of errors to be made. The technical error of measurement (TEM) was used, as it is the most commonly used measure of error in anthropometry (Mueller and Martorell 1988). The interobserver errors were calculated between the two measurers. Calculation of the interobserver errors automatically included the intra-observer errors, as the 2 people who took the measurements did not do so at the same time (Gordon et al. 2010). The formula for TEM is as follows:

$$\mathrm{TEM} = \sqrt{\left(\sum \mathrm{D}^2\right)/2\mathrm{N}}$$

Where D is the difference between two measurements of the same dimension of the same individual taken on two occasions and N is the number of individuals so measured (Stomfai et al. 2011; Perini e tal. 2005; Ulijaszek and Kerr 1999).

Comparisons of participants with themselves and others, were done using a modified Euclidean distance. In cases of all comparisons, differences between the two measurements of the same person, for all traits (either ratios or single dimensions) were divided by 2*TEM. Only integers of results of divisions were reported, thus any differences less than 2*TEM became zeros. This allowed the reporting of differences exceeding 95% confidence range of errors.

SPSS statistics and Microsoft Excel were used for all statistical analysis.

Results

Table 3 shows the means, standard deviations, coefficients of variation (CV) of the TEM and CV for individual dimensions, which were then used to calculate the ratios. There was a good range of variation within the sample with the largest standard deviation being for stature, which was 62.48 mm. The smallest TEM was for stature, it was 0.33% , which was 5.85mm. The greatest accuracy of measurements was 99.67%. The largest TEM was for the distance between the trochanterion to tibiale landmarks; it was 48.48mm or 11.08%. The largest CV TEM was 12% , which is for the distance between the subnasale and stomion landmarks. With the greatest TEM, the accuracy rate was still >88% for the most ideal conditions. This is comparable with other anthropometric studies[21].

Table 3: Means, SD, TEM, CV TEM and CV for individual dimensions for manual measurements.

	Mean	SD	TEM	CV TEM	CV		Mean	SD	TEM	CV TEM	CV
Ratio 1						**Ratio 6**					
a-ti	936.2	50.3	23.0	2.46	5.4	t-sy	709.4	49.6	17.6	2.5	7.0
tro-ti	437.5	45.0	48.5	11.08	10.3	sst-sy	527.6	44.1	22.8	4.2	8.4
tro-b	940.4	45.3	33.6	3.57	4.8	sst-b	1441.2	55.8	14.5	1.0	3.9
a-b	1439.0	61.8	6.1	0.42	4.3	t-b	1623.0	62.3	6.7	0.3	3.9
Ratio 2						**Ratio 7**					
m-sn	92.7	6.2	4.2	4.54	6.7	v-sst	315.5	19.3	17.3	5.5	5.9
n-sn	53.2	3.8	3.4	6.36	7.2	t-sst	181.7	18.9	15.5	8.4	10.3
n-sto	73.9	5.3	3.9	5.25	7.3	t-b	1623.0	62.3	6.7	0.5	3.8
m-sto	113.6	8.1	4.4	3.86	7.2	v-b	1756.7	62.5	5.8	0.2	3.6
Ratio 3						**Ratio 8**					
n-sto	73.9	5.3	3.9	5.25	7.3	a-ic (LHS)	363.6	30.6	9.8	2.7	8.3
sn-sto	20.7	3.4	2.5	12.00	16.5	ic-ic	314.5	37.6	12.3	3.8	12.0
sn-gn	65.3	4.6	3.1	4.77	7.1	a-ic (RHS)	363.6	30.6	9.8	2.7	8.3
n-gn	118.5	5.3	2.9	2.47	4.5	a-a	412.3	27.4	12.8	3.0	6.5
Ratio 4						**Ratio 9**					
a-ti	936.2	50.3	22.9	2.46	5.4	ic-zy (LHS)	227.5	20.2	5.8	2.6	8.8
sy-ti	410.6	36.5	33.8	8.24	8.9	zy-zy	141.2	5.3	2.5	1.8	3.6
sy-b	913.4	43.0	21.2	2.32	4.6	ic-zy (RHS)	227.5	20.2	5.8	2.6	8.8
a-b	1439.0	61.8	6.1	0.42	4.3	ic-ic	314.5	37.6	12.3	3.8	12.0
Ratio 5						**Ratio 10**					
v-sy	843.1	51.2	20.5	2.43	6.1	a-zy (LHS)	276.7	15.4	7.0	2.5	5.6
sst-sy	527.6	44.2	22.8	4.31	8.4	zy-zy	141.2	5.3	2.5	1.8	3.6
sst-b	1441.2	55.8	14.5	1.02	3.9	a-zy (RHS)	276.7	15.4	7.0	2.6	5.6
v-b	1756.7	62.5	5.8	0.33	3.6	a-a	412.3	27.4	12.8	3.0	6.5

Table 4 shows descriptive statistics of all ten ratios for all participants (n=20). The ratios have been calculated from the manual measurements, high quality photos and the low quality photos. The averages for each of the ratios were very similar between conditions, the largest difference was 0.53 for ratio 3, which was a difference of 30.7%, between the manual measurements and low quality photographs. Overall, the means did not differ significantly between ratios and conditions.

In the ideal condition, ie. ratios calculated from manual measurements, the TEMs were equal to or only slightly below the standard deviation (SD). The greatest difference between a TEM and a standard deviation for manual measurements was 0.03. Therefore, it can be seen that even in the most ideal conditions, measurement errors were responsible for majority of the variation of a ratio. In both high and low quality images, the TEM was equal to or larger than the standard deviation.

The CV TEMs shows the percentage of error for each ratio. There was a gradual increase in error between all three conditions eg. Manual measurements had the lowest error and low quality images had the highest error. This is to be expected. The largest consistent errors in all three conditions were seen in ratios 2 and 3 (Table 4).

Considering that the manual measurements have comparable TEMs, SD, means and CV with other studies (Gordon et al. 2013) (table 3), manual measurements or their derived ratios will not be discussed further.

Table 4: Means, SD, CV, R values, TEM and CV TEM for all ratios under all three conditions.

		Manual Measurements					**High quality photos**					**Low quality photos**				
	Landmarks	Mean	SD	CV - ratio	TEM	CV - TEM	Mean	SD	CV - ratio	TEM	CV - TEM	Mean	SD	CV- ratio	TEM	CV - TEM
Ratio 1	a,tro,ti,b	1.40	0.07	5.05	0.08	5.86	1.36	0.11	7.82	0.11	7.82	1.33	0.19	14.54	0.21	15.81
Ratio 2	m,n,sn,sto	1.14	0.03	2.79	0.03	2.40	1.17	0.09	8.01	0.10	8.01	1.22	0.18	14.58	0.16	12.88
Ratio 3	n,sn,sto,gn	1.99	0.19	9.65	0.16	7.93	1.88	0.33	17.61	0.31	17.61	1.46	0.28	19.06	0.27	18.30
Ratio 4	a,sy,ti,b	1.45	0.07	5.16	0.07	4.66	1.44	0.09	6.26	0.09	6.26	1.30	0.11	8.48	0.10	7.30
Ratio 5	v,sst,sy,b	1.31	0.04	2.72	0.03	2.06	1.27	0.04	3.27	0.04	3.27	1.24	0.03	3.16	0.04	3.25
Ratio 6	t,sst,sy,b	1.20	0.03	2.24	0.02	2.24	1.17	0.04	3.07	0.03	3.07	1.16	0.07	6.67	0.08	7.02
Ratio 7	v,t,sst,b	1.61	0.09	5.28	0.07	5.28	1.71	0.24	13.88	0.23	13.88	1.58	0.16	10.49	0.15	9.51
Ratio 8	a,ic,ic,a	1.02	0.02	1.58	0.01	1.58	1.01	0.02	1.87	0.01	1.87	1.16	0.21	18.26	0.18	16.00
Ratio 9	ic,zy,zy,ic	1.16	0.05	3.91	0.02	3.91	1.16	0.06	5.13	0.04	5.13	1.17	0.17	15.10	0.13	10.23
Ratio 10	a,zy,zy,a	1.32	0.03	2.52	0.02	2.52	1.27	0.09	7.27	0.10	7.27	1.33	0.15	11.25	0.14	10.83

An Euclidean distance was used to establish whether an individual could be correctly identified from the ratios calculated using low quality and high quality images, when the errors were taken into account. In an ideal situation, only the image of the participant that is compared with another image of the same participant should have no difference, i.e. 0, between measurements, after the TEMs were considered. Anything above a zero difference is concluded not to be a match for that person. Table 5 shows that individuals are being falsely identified as others, for example Participant 1(P1) and Participant 2 (P2) have a value of zero; this indicates that when errors are considered, there is no difference between them. At the same time, individuals who are themselves i.e. P1 and P1 are considered different, because there is a substantial difference between them in measurements from high quality and low quality images, which exceeds measurement errors. Participants were incorrectly identified 64.25% of the time.

Unlike a traditional Euclidean distance matrix, which is symmetrical, this one compares two different variables (i.e. high quality and low quality images), therefore a high quality image compared with a low quality image is above the diagonal, while below the diagonal is a low quality image compared with a high quality image. These variables were compared to establish whether the same person could be correctly identified from ratios when errors are considered and shown in images of different quality, as seen in real life forensic cases. This example shows that ratios do not have enough discriminatory power to differentiate between the correct and incorrect matches. This is further illustrated by the results seen in Table 4, which shows little difference between the average ratios for all face and body measurements

Table 5: A Euclidean distance between all males (n=20) of all ten ratios taken from high and low quality images by the same person. A measure of whether the ratios of each individual exceeds two times the TEM for that particular ratio.

		High quality																			
	Subject number	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**	**P11**	**P12**	**P13**	**P14**	**P15**	**P16**	**P17**	**P18**	**P19**	**P20**
Low quality	**P1**	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
	P2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0
	P3	1	1	1	1	1	1	0	1	1	1	0	0	0	1	1	2	1	0	0	1
	P4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
	P5	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
	P 6	0	0	1	0	1	1	1	1	1	0	1	1	1	0	1	2	1	1	1	1
	P7	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
	P8	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
	P9	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	1	1	1	1	0
	P10	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2	0	0	0	0
	P11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	1	0
	P12	1	0	0	0	1	0	0	0	1	1	1	1	0	0	1	2	1	1	1	1
	P13	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	P14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1
	P15	0	0	0	0	1	1	1	1	2	1	3	1	1	1	0	2	1	1	1	1
	P16	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	1	1	1	0
	P17	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0
	P18	0	1	0	0	1	0	0	0	0	1	1	0	1	1	1	1	1	0	0	1
	P19	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0
	P20	1	1	1	1	1	1	1	1	2	2	2	2	1	1	2	2	2	2	2	1

Note: A ‘0’ indicates a match, anything greater than or equal to 1 indicates a non match.

Individual dimensions were analysed, as the ratios did not have the power to discriminate between individuals (tables 4 and 5).The differences between measurements of the same individual taken by two people was analysed to examine whether it exceeds the TEM. Also, the measurements of each individual were compared with other individuals taking TEMs into account. High quality images were chosen for this exercise, as they represent the most optimal conditions for taking measurements from images. Table 6 shows that the number of traits (out of 40) differed by more than two times the TEM for each participant compared with themselves and all others, when the measurements were made from high quality photographs by two researchers. In reality, if an individual is compared with himself, there should be no differences which exceed the TEM, however, the findings presented in Table 6 contradicted this. The lowest difference between any participant and themselves was 2, i.e. 2 out of 40 traits differed by more than 2 times the TEM for that particular trait. This is not enough to make an identification of an individual. In some cases, the differences between a participant and themselves exceeded the differences between that particular participant and someone else, for example, P3 compared with P6, indicated that participant 3 and 6 are more likely to be identified as the same person than participant 3 is with himself. Results show that 34 out of the 40 traits did not exceed the acceptable error range. This was an accuracy of 85% once the errors were considered.

Even under optimal conditions, results show that measurements taken from images are not reliable, the question is, what traits can be reliably measured, if there are any?

Table 7 shows the number of times that a particular trait measured from high and low quality photographs differed by more than two times the TEM in the same individual. The trait with the largest number of differences was nasion-stomion length with 11 out of 20 differences, i.e. a difference of 55%. The trait with the lowest number of differences was the acromiale-zygomatic on the right side with 3 out of 20 differences, i.e. a difference of 15%.

Both high and low quality images had approximately the same total number of differences across all traits, high quality images had 119 differences, while low quality had 123 differences. Overall, results show that taking any measurement from any quality image is unreliable.

Table 6: The number of traits (out of 40) that differ by more than 2 times TEM for each participant compared with themselves and all others as measured from high quality photographs. This shows a comparison between participants as measured by TL (rows) and JK (columns).

Subject number	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**	**P11**	**P12**	**P13**	**P14**	**P15**	**P16**	**P17**	**P18**	**P19**	**P20**
P1	7	36	29	37	32	29	32	30	33	32	29	28	34	35	35	27	33	28	27	31
P2	34	7	19	7	12	13	14	18	5	18	20	25	16	9	12	16	1	21	25	10
P3	32	19	3	17	6	5	4	6	11	9	11	11	7	15	8	4	12	8	20	3
P4	32	20	8	17	1	10	3	14	18	10	12	19	4	8	10	10	13	10	23	2
P5	33	22	9	18	3	13	6	13	16	13	7	17	3	9	10	6	14	10	24	2
P6	29	22	1	22	6	4	5	8	14	11	16	17	6	19	6	9	17	9	23	10
P7	30	28	4	26	12	9	3	3	24	6	14	12	12	23	15	7	22	10	18	17
P8	29	26	7	25	11	12	5	8	21	10	14	11	13	18	17	4	22	10	19	12
P9	37	15	18	14	15	14	17	18	6	19	20	23	15	11	13	15	5	19	27	11
P10	33	22	10	20	8	5	5	4	13	5	15	12	14	13	11	11	13	17	20	9
P11	33	24	12	20	14	16	4	11	19	11	3	14	8	18	16	6	19	7	18	12
P12	33	26	17	24	20	17	13	11	23	10	16	6	20	24	22	16	21	20	15	20
P13	33	24	14	19	3	17	8	15	20	14	11	15	2	13	9	10	19	7	26	13
P14	35	21	18	18	9	12	14	18	16	17	23	21	18	8	15	22	14	21	27	9
P15	36	23	8	19	3	11	8	15	19	11	9	21	6	7	8	12	13	12	26	8
P16	32	23	9	21	6	10	1	6	19	5	10	13	9	16	11	5	15	11	18	10
P17	37	18	20	18	13	18	17	18	13	15	17	21	18	13	20	18	2	21	20	13
P18	32	22	8	20	8	12	1	9	19	11	4	11	5	16	13	4	16	9	18	9
P19	30	26	22	26	23	20	21	20	26	16	23	19	22	26	25	26	24	27	9	24
P20	36	21	11	17	2	12	12	14	18	14	9	17	10	6	13	15	13	12	20	4

Table 7: The number of times a dimension differed by more than two times TEM in an image of the same individual measured twice on high and low quality images.

	High quality (out of 20)	**Low quality (out of 20)**		**High quality (out of 20)**	**Low quality (out of 20)**
Ratio 1			**Ratio 6**		
a-ti	2	3	t-sy	3	3
tro-ti	4	5	sst-sy	2	4
tro-b	5	4	sst-b	4	3
a-b	2	4	t-b	3	4
Ratio 2			**Ratio 7**		
m-sn	3	2	v-sst	4	2
n-sn	3	4	t-sst	3	3
n-sto	7	4	t-b	3	4
m-sto	2	5	v-b	2	2
Ratio 3			**Ratio 8**		
n-sto	1	3	a-ic (L)	2	2
sn-sto	4	1	ic-ic	4	2
sn-gn	3	4	a-ic (R)	3	1
n-gn	4	2	a-a	3	2
Ratio 4			**Ratio 9**		
a-ti	2	3	ic-zy (L)	1	4
sy-ti	4	1	zy-zy	4	3
sy-b	4	4	ic-zy (R)	1	4
a-b	3	3	ic-ic	4	3
Ratio 5			**Ratio 10**		
v-sy	2	4	a-zy (L)	3	4
sst-sy	2	4	zy-zy	4	3
sst-b	4	3	a-zy (R)	1	2
v-b	2	2	a-a	2	3

Discussion

Previous studies which used ratios in an attempt to identify individuals (Kleinberg and Siebert 2012; Kleinberg, Vanezis and Burton 2007; Moreton and Morley 2011) have used only the face based on an empirically unsubstantiated belief that the face is the most variable part of the human body and thus is ideal for identification. Lucas and Henneberg (2015a) found that the body was in fact more variable than the face and is superior for identification purposes. Therefore, this paper investigated the use of ratios on the body and the face. The largest consistent errors between all three conditions were for the ratios that take measurements of the face (ratios 2 and 3). In both ratios, there was a difference of approximately 10% between manual measurements and low quality images. Details of the face are small and cannot be clearly seen in low quality images (Chen et al. 2010). Thus, the errors are higher as placement of anthropometric points accurately is difficult. On images, the head is one of the smallest parts of the body; therefore any inaccuracy in the placement of one point is reflected in the placement of all others as the relative distance between facial points is small. The smaller the distance, greater will be the chance of misplacing an anthropometric landmark. An example of this can be seen in Table 3, where the highest TEM is for the distance between the trochanterion and tibiale landmarks, which was 48.8mm, but the largest percentage of error was found when measuring the distance between the subnasale and stomion, which was only 2.5 mm.

Moreton and Morely (2011) state that proportions should only be used ‘to test for elimination’. The example presented in table 5 compares individuals with each other and themselves, and the individuals could not be excluded based on ratios. This is an important finding, as it illustrates the ineffectiveness of the proportions for any step in the identification of an individual. This is consistent with the findings reported by Moreton and Morely (2011). This was also tested on single dimensions and results were similar (table 6).

Porter and Doran (2000) have claimed that only horizontal proportions should be used, as the proportions in the vertical plane undergo image distortion. Although, this was tested on facial proportions, the same should hold true for the body proportions. Moreton and Morley (2011) showed that all proportions vertical or horizontal were affected by image distortions. However, the use of the anharmonic ratio in the current study, should remove any effects of angular and rectilinear image distortions. Even though, there was no image distortion, findings of the current study indicated that all anharmonic ratios whether taken from the face, body, horizontally or vertically using high or low quality pictures, there was an error range (Table 4) that deems their use unreliable for forensic purposes.

The anharmonic ratio uses a combination of four measurements. Each of those measurements is taken with an error, even in cases where the error is small i.e. within the accepted error range, as seen in the manual measurements (Table 3), the error of the ratio is approximately the same as the standard deviation. When measurements with an error are combined, their errors are also combined. Thus, any variation in the ratio is due to errors and not actual human biological variation. This could result in someone being identified as someone else (Table 5). In real forensic cases, this outcome would be disastrous. This makes ratios an unacceptable tool for the identification of individuals from images.

Conclusion

The error rates of dimensions measured from images increase as the quality of the images decreases. Due to the high error rates, ratios are not considered as a reliable method for the identification of individuals. This paper shows the error rates for ratios and single dimension measurements taken from high and low quality images, in both cases the errors exceed those of manual measurements. Therefore, taking measurements from images generate high error rates and make them unreliable in the identification of an individual, particularly in the court of law. Furthermore, they do not meet the Daubert Criteria.

Acknowledgements

The authors would like to thank Professor John Fryer for providing his expertise in relation to anharmonic ratio. A very special thank you to Tavik Morgenstern for his assistance with photography and figure design.

References

Australian Bureau of Statistics 2013, ‘Gender indicators, Australia’, cat. no. 4125.0, ABS, Canberra.

BenAbdelkader, C, Davis L 2006 ‘Estimation of anthropomeasures from a single calibrated camera’, in: *Proceedings of International Conference on Automatic Face and Gesture Recognition*, Southhampton pp. 499–504.

BenAbdelkader, C, Yacoob, Y, 2008 ‘Statistical body height estimation from a single image’, in: *Proceedings of International Conference on Automatic Face and Gesture Recognition*, Amsterdam, pp. 1–7.

Biber, K 2009, ‘Visual jurisprudence: the dangers of photographic identification evidence’. *Criminal Justice Matters*, vol.78, pp. 35-37.

Chen, S, Mau, S, Harandi, M, Sanderson, C, Bigdeli, A, Lovell, B 2010, ‘Face recognition from still images to video sequences; A local-feature-based framework’, *Journal on Image and Video Processing*, vol. 2011, pp. 1-14.

Crumbley, L, Cheng, C 2014, ‘Avoid losing a Daubert challenge: some best practices for expert witnesses’. *The ATA Journal of Legal Tax Research*, vol. 12, pp. 41-53.

Edmond, G 2008, ‘Specialised knowledge, the exclusionary discretions and reliability: Reassessing incriminating expert opinion evidence’. *University of New South Wales Law Journal*, vol. 31, pp.1-55.

Edmond, G 2010, ‘Impartiality, efficiency or reliability? A critical response to expert evidence law and procedure in Australia’, *Australian Journal of Forensic Science*, vol. 42, pp. 83-99.

Edmond, G, Biber, K, Kemp, R, Porter, G 2009, ‘Law's looking glass: expert identification evidence derived from photographic and video images’. *Current Issues in Criminal Justice*, vol. 20, pp. 337-376.

Fryer, JG 2000, ‘An object space technique independent of lens distortion for forensic videogammetry’. *International archives videogammetry Remote Sensing*, vol. 33, pp. 246-252.

Gordon, C, Claire, C, Blackwel,l CL, Bradtmiller, B, Parham, JL, Hotzman, J, Paquette, SP, Corner, BD, Hodge, BM 2010, *Anthropometric Survey of US Marine Corps Personnel: Methods and Summary Statistics*. No. NATICK/TR-13/018. Army Natick Soldier Research Development and Engineering Center MA, 2013.

Grivas, CR, Komar, DA 2008, ‘Kumho, Daubert and the nature of scientific inquiry: implications for forensic anthropology’. *Journal of Forensic Sciences*, vol. 53, p.771-766.

Ireland, JL, Beaumont, J 2015, ‘Admitting scientific expert evidence in the UK: reliability challenges and the need for revised criteria – proposing an Abridged Daubert’. *Journal of Forensic Practice*, vol. 17, pp. 3-12.

Kleinberg, KF, Siebert, JP 2012, ‘A study of quantitative comparisons of photographs and video images based on landmark derived feature vectors’, *Forensic Science International*, vol. 219, pp.248-258.

Kleinberg, KF, Vanezis, P, Burton, M 2007, ‘Failure of anthropometry as a facial identification technique using high-quality photographs’, *Journal of Forensic Science*, vol. 52, pp.770-783.

Lucas T, Henneberg, M 2015a, ‘Comparing the face to the body, which is better for identification?’, *International Journal of Legal Medicine*. DOI: 10.1007/s00414-015-1158-6.

Martin, R, Saller, K 1957 *Lehrbuch der antropologie*. Stuttgart, Gustav Fischer Verlag.

Moreton, R, Morley, J 2011, ‘Investigation into the use of photoanthropometry in facial image comparison’, *Forensic Science International*, vol.212, p.231-237.

Mueller, WH, Martorell, R 1988, *Reliability and accuracy of measurement. In Anthropometric Standardisation Reference Manual*, Champaign, IL: Human Kinetics Books.

Perini, TA, Oliveira, GLD, Ornellas, JDS, Oliveira, FPD 2005, ‘Technical error of measurement in anthropometry’, *Revista Brasileira De Medicina Do Esporte*, vol. 11, pp. 86-90.

Porter, G, Doran, G 2000, ‘An anatomical and photographic technique for forensic facial identification, *Forensic Science International*, vol. 114, pp.97-105.

Scoleri, T, Henneberg, M 2012, ‘View-independent prediction of body dimensions in crowded environments’. *International Conference on Digital Image Computing Techniques and Applications (DICTA)*; 3-5 December 2012; Fremantle.

Scoleri, T, Lucas, T, Henneberg, M 2014, ‘Effect of garments on photoanthropometry of body parts: Application of stature estimation’, *Forensic Science International*, vol. 237, pp.1-12.

Stomfai, S, Ahrens, W, Bammann, K, Kovács, E, Mårild, S, Michels, N, Moreno, LA, Pohlabeln, H, Siani, A, Tornaritis, M, Veidebaums, T, Molnár, D 2011, ‘Intra- and inter-observer reliability in anthropometric measurements in children’, *International Journal of Obesity*, vol. 35, pp. 545-551.

Ulijaszek, SJ, Kerr, DA 1999, ‘Anthropometric measurement error and the assessment of nutritional status’, *British Journal of Nutrition*, vol. 82, pp. 165-177.

Discussion

The manuscripts in this thesis all have a common theme, establishing the usefulness and reliability of taking measurements of the human body to be used for image identification.

It was established that interval scale measurements of the human body could be taken from images (manuscript 1). Body dimensions can be measured directly, or, be predicted based on others (authors used upper limb length to predict stature). Prediction of dimensions has particular significance in real life forensic cases when the entire body cannot be seen on an image. The reliability of these measurements was discussed using bias and technical error of measurements. Measurement errors were calculated between individuals of different experience levels in locating anthropometric points on the human body. A 95% accuracy was reported when estimating stature based on upper limb measurements. Although the accuracy rate is impressive and accepted scientifically, in reality, it would not be sufficient for identification from images. An inaccuracy of 5% when applied to the average stature can be calculated at approximately 80mm. If these errors are taken into account and considered, the suspect pool would be widened significantly (manuscript 4). Reported error rates are not optimal for real life forensic cases. Even if more measurements are used in an attempt to isolate an individual (as suggested in manuscript 4), if all measurements are taken with the same error the probabilities associated with achieving singularity would be minimal.

Whilst testing the accuracy of taking measurements from images (manuscript 1) it was suggested that knowledge of the human body and training in locating anthropometric points, improves results. This finding is further substantiated with the results presented in manuscript 5 which show minimal measurement errors between three trained anatomists. Although the methods in these two manuscripts are different, the finding remains the same. This finding illustrates the need for biological anthropologists to continue their role in forensic identification from images.

Different types of clothing were used to investigate their effect on the placement of anthropometric points (manuscript 1). It was found that the black shirt had the lowest error, the paper suggested that this was because of gravity acting in a way that allows the acromiale to easily be observed. However it has also been suggested that assessors were inaccurate with garments other than the black shirt because they were using the seam of the sleeve to indicate the position of acromiale. The seam is not visible in the black shirt which produced more accurate results. This point warrants discussion, as the leather jacket was a single dark colour like the black shirt it is unlikely that assessors were using other cues in their location of points. However this is an aspect of that study which could benefit from further research which studies outside cues in identifying anthropometric points. The results of this paper found that thicker garments produced higher inaccuracies. However, with repetition these errors were minimised which further supports the role of experienced anatomists in identification from images. The findings from manuscript 1 directly impacted the method in manuscript 5. Participants were imaged using CCTV surveillance while clothed, as it was found that clothing had little impact when persons with anatomical knowledge were placing points on the body.

This thesis answered the question, ‘how useful is anthropometry of the entire human body in the identification of individuals’. Only the use of craniometry in the identification of individuals has been previously studied (Schimmler, Helmer and Rieger 1993). In a study by Schimmler and colleagues (1993) a complex method was used to analyse the individuality of craniometric measurements. The authors concluded that each individual had their own set of craniometric measurements which did not match those of any other in the study. However ,this study only used a sample of 95 skulls. The study by Schimmler and colleagues (1993) presented a very complex method to analyse individuality. The overall goal of anyone researching forensic methods should be to present a method which can easily be understood and thus used in court proceedings. A study by Goldstein (1971) used categories to describe the individuality of the human face, this study only used a sample of 256 faces. Much like the study by Schimmler (1993), Goldstein did not take into consideration correlations between traits and used a small sample. Manuscript 2 presented a simple method which can be applied to large databases of human measurements. The Interval scale of measurement was chosen because it is deemed more reliable in court proceedings.

No study which has assessed the usefulness of anthropometry in isolating an individual from a sample has used a combination of face and body measurements, nor compared the two. A recent literature review was conducted by Gibelli et al. (2016) which show that the only anthropometric dimension used in image analysis is base to vertex (height). Measurements of the body have not previously been used in isolation studies. Manuscript 3 assessed which measurements are better for identification (the face or the body), it ultimately concluded that the body is better at isolating an individual.

Although neither manuscript 2 or 3 o used measurements taken from images, it is a significant finding for image identification. It can assist future researchers by identifying the parts of a human with the highest identification rate which can then be tested using image identification technologies. Although both manuscripts 2 and 3 were successful in achieving singularity using eight or less measurements, the smallest amount of measurements needed is preferable. Manuscript 1 presented a number of images where the full body was not visible, this is a true representation of the types of images captured using CCTV. The theoretical findings of manuscripts 2 and 3 are extremely useful as not all parts of the human body can be seen on an image and thus measured. Therefore priority of measurements should be ranked from largest to smallest.

The overall findings of the first four manuscripts are that the human body is an extremely useful tool for identification; however its usefulness as a biometric tool depends on the accuracy of technology which can measure distances from images. These technologies are still being developed and are not sufficient for real life forensic identification. Therefore the authors attempted to bypass the errors associated with technology by introducing a simple mathematical ration to assess the human body. Manuscript 5 failed to produce a reliable method which can be used to identify individuals from images, even when using the face and body. Even though all of the assessors in manuscript 5 were trained anatomists which produced smaller errors, the method of applying ratios could not achieve singularity. This is because ratios of body dimensions (proportions) vary less than sizes of bodies, while errors of individual measurements become compounded when ratios are calculated. Thus less variation and bigger errors produce less opportunity for discrimination.

In summary, it can be suggested that there needs to be a further collaboration between biological anthropologists and computer analysis experts. Knowledge of the human body and its usefulness as a biometric tool has helped to develop the methods for image identification. However more work needs to be done to minimise the errors produced by measurement technologies.

References - specifically in discussion

Gibelli, D, Obertova, Z, Ritz-Timme, S, Gabriel, P, Arent, T, Ratnayake, M, De Angelis, D, Cattaneo, C 2016, ‘The identification of living persons from images: A literature Review’, Legal Medicine, vol. 19, pp. 52-60.

Goldstein, AJ, Harmon, LD, Lesk, AB 1971, ‘Identification of human faces’, *Proceedings of the IEEE*, vol.59, no. 5, pp.748-760.

Schimmler, JB, Helmer, RP, Rieger, J 1993, ‘craniometric individuality of Human Skulls’, in Isçan MY, Helmer, R (ed.), *Forensic analysis of the skull*, Wiley-Liss, New York, USA.

Appendix 1 : Reprint of ‘Effect of garments on photoanthropometry of body parts: Application of stature estimation’.

Forensic Science International 237 (2014) 148.e1–148.e12

Contents lists available at ScienceDirect

Forensic Science International

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

Forensic Anthropology Population Data

Effects of garments on photoanthropometry of body parts: Application to stature estimation

Tony Scoleri [a],*, Teghan Lucas [b], Maciej Henneberg [b]

[a] *Image Analysis and Exploitation Group, Defence Science and Technology Organisation, PO Box 1500, Edinburgh 5111, Australia*
[b] *Biological Anthropology and Comparative Anatomy Research Unit, School of Medical Sciences, University of Adelaide, Adelaide 5005, Australia*

ARTICLE INFO

Article history:
Received 30 August 2013
Received in revised form 25 October 2013
Accepted 30 December 2013
Available online 13 January 2014

Keywords:
Human height estimation
Person identification
Human characterisation
Image measurement
Clothing effects
CCTV

ABSTRACT

Person identification from images is an important task in many security applications and forensic investigations. The essence of the problem comes down to measuring key observable anatomical features which can help describing similarities or differences between two or more individuals. In this paper, we examine how different types of garments affect the placement of body markers that enable precise anatomical human description. We focus in particular on landmark positioning errors on the upper limb. Closed-form formulae are provided to compute the maximum likelihood estimate of upper limb length from an image. Subject stature is then predicted from the limb length through a regression model and used as identification criterion. Following initial laboratory experiments, the technique is demonstrated to be invariant to posture and applicable to uninformed subjects in unconstrained environments. Seven technical errors of measurement and statistical tests are quantified empirically from statures obtained by three assessors. Results show that thicker garments produce higher inaccuracies in landmark localisation but errors decrease as placement is repeated, as expected. Overall, comparison to truth reveals that on average statures are predicted with accuracy in excess of 96% for the worst assessor.

1. Introduction

The use of closed circuit television videos (CCTV) in crime control and prevention has rapidly grown in the present security context. As a counteraction, criminals often disguise their faces to preclude identification. When a person lies within close proximity of a camera, their facial features can be acquired with high fidelity. Reliable identification is then achievable, for instance, by finding the smallest distance between the outline of anatomical landmarks on the target face and the corresponding outline of 3-D faces in a gallery [1]. This assumes that the target face is only partially masked and, depending on the viewing geometry and visible facial components, different anatomical points must be identified every time. Further manual intervention is required to register (orientate and scale) the outline of the target face to that of each 3-D face. This task is extremely labour intensive and time consuming the larger the database is.

For the majority of surveillance systems though, the quality of the recorded imagery remains poor. The cost associated with better optical and digital components is often considered too high so many institutions opt for quantity rather than quality [2]. Recorded CCTV images thus display optical distortions, object blur, wrong colours and have low resolution which makes fine details invisible [3,4]. To circumvent these issues, attempts have been made using anthropometry to identify individuals based on larger and more easily observable aspects, such as body shape [5,6]. The use of anthropometric measurements have been included in the description and identification of individuals from photographs as early as the 19th century [7,8] and today they are combined with image analytical techniques [9–13].

When looking at body shape, the choice of garment has been reported to alter the perception of the human body [14]. In particular, the colour of clothing is noted to have an effect on body appearance. In a survey of male subjects [15], it is reported that black is chosen when trying to minimise body size, and light colours, such as white, are worn to maximise body shape definition. The optical illusions presented by patterns such as horizontal stripes are considered to have a contradictory effect by either making the body appear wider and shorter [16] or slimmer and taller [17]. The size of clothing can also affect an individual's perception of body shape, presenting the wearer as larger or smaller than normal depending on the fit of clothing to their actual body parts [18]. Latest research using LIDAR technology [19] has demonstrated that gender classification can be achieved with an accuracy of 70–80% for subjects wearing three clothing styles

---

* Corresponding author. Tel.: +61 8 7389 7946.
E-mail addresses: tony.scoleri@dsto.defence.gov.au (T. Scoleri), teghan.lucas@adelaide.edu.au (T. Lucas), maciej.henneberg@adelaide.edu.au (M. Henneberg).

0379-0738/$ – see front matter. Crown Copyright © 2014 Published by Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.forsciint.2013.12.038

(summer, fall, winter). Another recent study [20] has reached similar conclusions, that clothed individuals seen in CCTV images can only be matched above random expectation for a general body shape. These two lines of research have proved independently that the acquisition of body descriptors from images is largely deceived by clothing whether it is obtained from automatic machine learning or based on human perception. Although this statement is intuitive, the challenge is to acquire those descriptors with highest precision from a range of scenarios and body postures.

Our criterion for person identification relies on estimating the body height or stature. When the entire body is visible, image metrology techniques and virtual scene superimposition can be used to measure stature directly from the top of the head to the feet [21–23]. In urban scenes though, pedestrians are often partially occluded in a way that prevents application of such techniques. In other circumstances, only a single image of the person is available rather than a video so stature cannot be derived from gait analysis. Recent research has thus looked at reconstructing the total body height from body parts measurements [13,24]. Clothing style then affects the accuracy to which this may be done. The work in this paper quantifies the effects of garments on photoanthropometry of body parts by first measuring the upper limb length of a person, predicting their stature by linear regression from this length and finally assessing errors from the obtained stature.

In all error functions, part of the error is induced by either the human intervention, when manually locating anthropometric points, or by automatic body part detectors and classifiers. This paper focuses on assessing the human error in the process. Initial laboratory experiments have been conducted on nine male participants wearing no shirt, a black shirt, a horizontally striped shirt and a padded leather jacket. Each of these garments is specifically chosen to evaluate how garment colour, pattern and type alter an assessor's perception on body landmark location. Subsequent experiments estimate the errors for eighteen uninformed male subjects observed in an uncontrolled (airport) environment. Three assessors have examined the videos and marked points to measure the upper limb length of each subject. The assessors come with varying knowledge in anatomy and computer vision which provides valuable feedback on the spread of errors.

The main contribution of this research is to quantify the variations in placing critical body landmarks when these points are covered by different garment styles (Section 3). Seven error measures are described to estimate those variations and their statistical significance. The second contribution is the overall framework (Section 2) which includes robust image analytical techniques. In particular, closed-form formulae are proposed for computing the maximum likelihood estimate of the upper limb length (Section 2.2). This framework is advantageous especially for its application to uncontrolled scenes, independence from subject posture and flexibility of integrating other body part measurements as necessary. Contrary to many identification techniques based on facial recognition, the proposed method can be applied to low resolution images and offers limited user interaction (only a few clicks are necessary). This comes as a result of the particular choice of anthropometric landmarks which are distinguishable under arbitrary perspective camera views and suppress the need for intensive image manipulation or alignment. The achievement is that, for stature estimation, average accuracy (or recognition rate) is in excess of 96% for the worst assessor when compared to truth.

2. Stature estimation

For some years now, researchers in computer vision have tried to find appropriate markers for soft biometry retrieval from videos. Most notable is the work on estimating a person's stature. When the scene is accessible for surveying, a virtual model can be created and superimposed onto the original imagery [21]. As for face recognition, human stature can accurately be measured when subjects walk in the vicinity of the camera. In general though, alternative methods are needed in cases where the environment is more challenging (e.g. outdoors), surveying is not possible (e.g. hazardous trafficking area) or subjects are remote from the camera and not standing in perfectly upright position. Besides, whether the person appears in a single frame or is seen in motion through a video sequence, great difficulties arise in precisely extracting the top of the head (vertex) and heel position on the ground [24–28]. In all undertakings, the process is image-driven, relying on some variant of silhouette extraction to define the head and feet locations. This resulted in different authors having different definitions for how these points should be retrieved from images. Recent research [24] has settled some of these questions, however, when measuring several partial body dimensions to obtain complete stature, the location of body part markers remains debatable.

In contrast, the scheme we propose is evolved from a well-defined anatomical model of body landmarks extensively tested in anthropological research. The uncertainty in extracting the landmarks from images becomes of prime importance before carrying out any measurement. This is examined in our experiments as well as its effect on a person's stature estimation. Forensic scientists and anthropologists routinely perform measurements of long bones to reconstruct total body height from regression equations that relate those body parts to the human stature [29,30]. Such reconstructions have been widely successful and achieve about 95% accuracy on body height prediction [31]. Statistically, the upper limb length relates significantly to the human stature and is commonly observable and measurable in CCTV videos. Besides, its length does not undergo any diurnal change unlike stature. It is indeed a well-recognised phenomenon that stature begins to decrease immediately after rising in the morning and further loss continues throughout the day up to a maximum of 28.1 mm [32]. A person

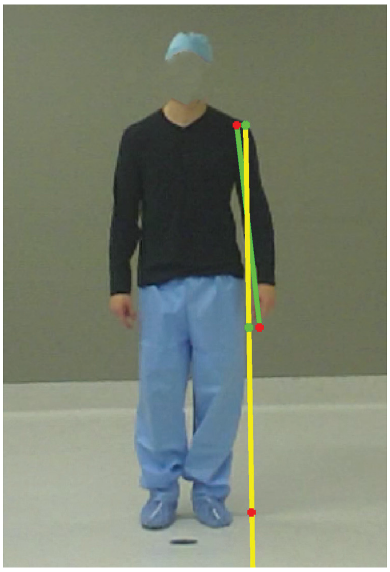


Fig. 1. Upper limb model. The user-marked points are shown in red and computed ML points in green.

should be measured preferably in the afternoon to reduce the variation in stature as loss of height occurs most rapidly in the morning [32,33]. Since this constraint is not realisable in general surveillance context, the upper limb length presents a good alternative to infer stature.

2.1. Upper limb model

The upper limb length of a subject is measured as

$$u = \operatorname{sgn}(h_a - h_d) \cdot (h_a - h_d),$$

where sgn (x) stands for the signum function of a real number x, h_a is the height from a ground point to the acromiale (the shoulder point) and h_d is the height from the same ground point to the dactylion (the tip of the middle finger). These heights are obtained from an image by asking a user to place three markers corresponding to the acromiale, dactylion and a point on the ground. The use of the signum function enables to mark the acromiale and dactylion in any order of preference. One assumption here is that these two points are situated at the same depth in the scene and therefore define an upright vertical segment to the ground. Their depth is estimated as the distance from the third point placed on the ground and a world origin.[1] Once the acromiale and dactylion points are set, their corresponding Maximum Likelihood (ML) location is computed such that the ML points are aligned with the vertical scene direction, see Fig. 1. Details of the alignment procedure are deferred to the next section.

To assist the user in choosing the ground point, a guide line is drawn through the ML points. The ground marker can then be set along the line where the heels of the subject are touching the floor. A short procedure is also applied to ensure that this point lies on the guide line perfectly. With the three collinear points, heights h_a and h_d are orthogonal to the ground and can be calculated as described in Section 2.3.

2.2. Maximum Likelihood estimation of limb endpoints

Image perspective, subject posture and user interaction mean that the upper limb segment will rarely stand in the vertical scene direction required to measure its length. This limitation is addressed by computing new endpoints which are aligned with the vertical vanishing point, $\mathbf{v}_z$, obtained during calibration.

Suppose that the input markers are given by two points $\mathbf{x} = [x_1, x_2]^{\mathrm{T}}$ and $\mathbf{x}' = [x'_1, x'_2]^{\mathrm{T}}$ with associated 2×2 isotropic Cartesian covariance matrices $\Lambda_{\mathbf{x}}$ and $\Lambda_{\mathbf{x}'}$ defining perturbation around $\mathbf{x}$ and $\mathbf{x}'$ by circles of radius r and r', respectively. The Maximum Likelihood estimates $\bar{\mathbf{x}}$ and $\bar{\mathbf{x}}'$ of input markers $\mathbf{x}$ and $\mathbf{x}'$ can be determined by minimising the squared Mahalanobis distance

$$d^2_{\mathrm{Mahal}} = (\mathbf{x} - \bar{\mathbf{x}})^T \Lambda_{\mathbf{x}}^{-1} (\mathbf{x} - \bar{\mathbf{x}}) + (\mathbf{x}' - \bar{\mathbf{x}}')^T \Lambda_{\mathbf{x}'}^{-1} (\mathbf{x}' - \bar{\mathbf{x}}')$$

subject to the alignment constraint $v_z^T \boldsymbol{l} = 0$, with $\boldsymbol{l} = [\bar{\mathbf{x}}^T, 1]^T \times [\bar{\mathbf{x}}'^T, 1]^T$. This is a constrained optimisation problem which can be solved in closed-form using the Lagrange multiplier method. For $v_z = [v_1, v_2, v_3]^T$ and $\mathbf{z} = [r, r', \mathbf{x}^T, \mathbf{x}'^T, v_z^T]^T$, it can be shown that

$$\boldsymbol{l} = \begin{bmatrix} 1 + \sqrt{1 + [\xi(\mathbf{z})]^2} \\ \xi(\mathbf{z}) \\ -v_1 v_3^{-1}\left(1 + \sqrt{1 + [\xi(\mathbf{z})]^2}\right) - v_2 v_3^{-1} \xi(\mathbf{z}) \end{bmatrix},$$

where the real-valued rational function $\xi : \mathbb{R}^9 \mapsto \mathbb{R}$ has the form

$$\xi(\mathbf{z}) = 2\,\frac{r' d_1 d_2 + r d'_1 d'_2}{r'(d_1^2 - d_2^2) + r(d'_1 2 - d'_2 2)}$$

with $d_i = x_i - v_i v_3^{-1},\ d'_i = x'_i - v_i v_3^{-1},\ i = 1, 2$. The previous formulae hold as long as the vertical vanishing point is not ideal ($v_3 \neq 0$).

In our implementation, anisotropic Cartesian covariances $\tilde{\Lambda}_{\mathbf{x}}$ and $\tilde{\Lambda}_{\mathbf{x}'}$ are first calculated [34] and then employed to yield

$$r = |det(\tilde{\Lambda}_{\mathbf{x}})|^{1/4} \quad r' = |det(\tilde{\Lambda}_{\mathbf{x}'})|^{1/4}.$$

Writing $\boldsymbol{l} = [l_x, l_y, l_w]^T$, the ML estimates of $\mathbf{x}$ and $\mathbf{x}'$ are given by the Cartesian coordinates

$$\bar{\mathbf{x}} = \left[\frac{x_1 l_y^2 - x_2 l_x l_y - l_x l_w}{l_x^2 + l_y^2}, \frac{x_2 l_x^2 - x_1 l_x l_y - l_y l_w}{l_x^2 + l_y^2}\right]^T,$$

$$\bar{\mathbf{x}}' = \left[\frac{x'_1 l_y^2 - x'_2 l_x l_y - l_x l_w}{l_x^2 + l_y^2}, \frac{x'_2 l_x^2 - x'_1 l_x l_y - l_y l_w}{l_x^2 + l_y^2}\right]^T.$$

These points are taken as the true locations of the upper limb endpoints. The above derivation differs from its original form in [22] in that critical entities are readily programmable as given here with some of them explicitly calculated in Cartesian rather than projective coordinates to prevent potential errors.

2.3. Upper limb length

Without loss of generality, suppose we wish to determine the actual height h_a from the ground point **G** to the acromiale **A**. Let **g** and **a** denote their corresponding image points, with **a** the ML estimate of the user-defined acromiale. Assuming a perspective projection camera model, these relationships may be written as

$$\left\{ \lambda_1 [\mathbf{a}^T, 1]^T = P\,[\mathbf{A}^T, 1]^T, \lambda_2 [\mathbf{g}^T, 1]^T = P\,[\mathbf{G}^T, 1]^T, \right.$$

where $\mathtt{P}$ encodes the projection matrix and the λ_i's some perspective scale factors. The above system of equations can be expressed in matrix form as $\mathtt{M}\mathbf{A} = \mathbf{b}$. This follows from using the assumption that **A** and **G** are at the same depth, so one may write $\mathbf{A} = [A_1, A_2, h_a]^{\mathrm{T}}$ and $\mathbf{G} = [A_1, A_2, 0]^{\mathrm{T}}$ which provides four equations in three unknowns. The least-squares solution $\widehat{\mathbf{A}} = [M^T M]^{-1} M^T \mathbf{b}$ gives h_a as the third component of $\widehat{\mathbf{A}}$.

Height h_d can be calculated in a similar manner. The upper limb length ensues from the formula given in Section 2.1.

2.4. Anthropometric stature prediction

Anthropometric data were collected from 109 adult males resident in Australia. These included upper limb length and body height measured in accordance with the Martin's Technique [35] and in compliance with the International Standards Organisation (ISO 7250). Two linear regression techniques, namely the Ordinary Least-Squares (OLS) and Reduced Major Axis[2] (RMA), can be applied to these data to predict stature from upper limb length.

---

[1] The world origin is chosen on the ground during camera calibration.

[2] RMA is also known as the Total Least-Squares method.

Fig. 2. Anterior views of the same male in four different types of wear. (a) Shirtless; (b) black shirt; (c) striped shirt; (d) padded jacket.

If u and s denote the upper limb length and stature respectively, both expressed in millimetres (mm), then

$$\begin{aligned}\text{OLS}: &\quad s = 1.4052u + 678.74,\\ \text{RMA}: &\quad s = 1.7435u + 413.94.\end{aligned}$$

The variances associated with the OLS and RMA stature predictions are 48.5 and 50.6 mm, respectively.

3. Experiments

Clothing effects are first examined under controlled conditions in a laboratory (Sections 3.1–3.3). Various qualitative measures are calculated to evaluate the errors in landmark placement and differences between assessors. Visual influence of garment styles is deduced from those errors. Section 3.4 reports the results when our technique is applied to uninformed subjects in an unconstrained (airport) scene. In all laboratory tests, only the OLS regressor is used to predict stature from upper limb length. The RMA method is temporarily discarded because RMA statures are linearly related to OLS ones and therefore would reveal similar trends. RMA is used in Section 3.4 on every-day life surveillance imagery taken in an airport.

3.1. Laboratory experiments overview

3.1.1. Laboratory set-up

Nine adult male participants have been recruited within South Australia. Each of them is made to wear a pair of surgical pants, shoe coverings, a cap and a face mask to eliminate identifying features. This strategy also intends to focus the assessors' attention on the upper body with no other distractions. Participants are recorded using a CCTV camera (axis p3304 with a resolution of 1280 × 800 pixels) in the Bioskills laboratory of the Medical School at The University of Adelaide. The camera is fixed to the ceiling at a height of 2.5 m from the floor. Still photographs are extracted from the videos and calibrated using the technique in [36]. Each participant is standing approximately 8 m from the camera shown in either an anterior or posterior view. They are imaged four times wearing no shirt, a black shirt, a horizontally striped shirt and a padded leather jacket ($n = 36$ photographs). Fig. 2 shows the garments on a participant. In order to reduce the influence of diurnal variation on stature, the men have been measured and recorded in the afternoon. This provided an estimate of their "true" stature. The quotation marks are used because although every care has been taken to minimise the errors in true statures, it was reported that some may still exist [37].

Several challenges became apparent when analysing the images. Despite instructing the participants on the correct pose to hold, they are often slouching or leaning to one side, their upper limb is slightly bent and not straightened out, their hands are curled up and not fully opened, their feet are not together but separated (Fig. 3). These various postures do not adhere to the correct anatomical model and therefore introduce errors. However, they offer realistic conditions as would be encountered in real-life CCTV images, which is important.

3.1.2. Assessors

Three assessors have viewed each of the photographs and marked the three points as described in Section 2.1. Their level of experience in anatomy can be ranked as expert, trained and novice. The "expert" has over forty years of experience studying and measuring the human body and is employed as a Professor of Anatomical Sciences, teaching students as well as conducting research in the field. The person who is referred to as "trained" is educated in the field of Biological Anthropology and has three years of experience in measuring and studying the human body. The "novice" assessor, although expert in computer vision gait analysis, has limited experience in defining a person's anatomical features. The broad range of expertise is valuable for observing the variations of errors.

In the experiments, assessors only have a single attempt at marking points for each participant. This guarantees integrity in revealing the effects of garments on landmark placement. It also means that error values can improve if an assessor could mark points repeatedly. Assessor 2 was chosen to complete the task twice in order to calculate the intra-observer error. The landmark positioning took just over an hour for each assessor to go through the 36 images; assessor 2 was given a two-hour break in between

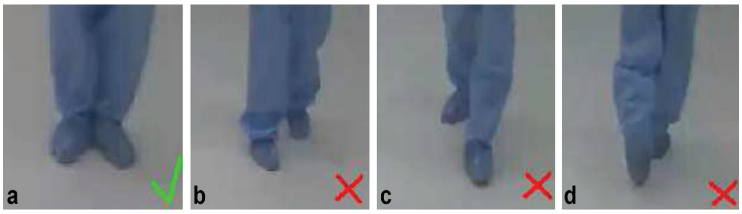


Fig. 3. Sample images of feet position for different participants. (a) Correct anatomical position; (b–d) incorrect pose.

repeats in order to reduce the effects of fatigue and memory on the task.

3.2. Error analyses for three uncertain markers

The Technical Error of Measurement (TEM) is a suitable quality measure to assess the difference between stature measurements [37]. Several such TEMs are described in the following sections. In addition, the bias between assessors' measurements and the significance of variances between TEM values are examined. In total, seven tests are presented to understand how errors fluctuate under the influence of landmark positioning and garment style.

3.2.1. Comparison to truth

The first TEM evaluates the discrepancy between all predicted statures and their "truth" values. It is given by the formula

$$\mathrm{TEM}_{\mathrm{truth}} = \left(\frac{1}{2n}\sum_{i=1}^{n}(\bar{s}_i - s_i)^2\right)^{1/2},$$

where n is the total number of test images, $\bar{s}_i$ and s_i stand for the true and predicted statures, respectively. The assessors produced TEMs as shown in Table 1.

In our model, the total error on stature prediction stems from a combination of the errors in placing the anthropometric points, calibrating the camera and the OLS regressor. Given that the variance for OLS is estimated at 48.5 mm, these errors are within that threshold and thus very encouraging. Considering the worst score and comparing to the average participant's height of 1758 mm yields an accuracy in excess of 97%. For the best TEM score, the accuracy reaches over 98%.

3.2.2. Inter-observer error

An inter-TEM is used to measure the extent to which predictions from assessors differ from one another:

$$\mathrm{TEM}_{\mathrm{inter}} = \left(\frac{1}{n(k-1)}\sum_{i=1}^{n}\left[\sum_{j=1}^{k}s_{i,j}^2 - \frac{\left(\sum_{j=1}^{k}s_{i,j}\right)^2}{k}\right]\right)^{1/2},$$

where $s_{i,j}$ is the i-th predicted stature obtained by the j-th assessor and k refers to the total number of assessors (here $k = 3$). Table 2 shows the values obtained for different selections of assessors.

More experienced anatomists (assessors 1–2) have a lower TEM than less experienced anatomists (assessors 2–3). So, these results confirm the assessor's experience in anatomy.

Table 1
Errors (mm) between predicted statures and truth.

	Assessor 1	Assessor 2	Assessor 3
$\mathrm{TEM}_{\mathrm{truth}}$	30.0	39.2	44.3

Table 2
Inter-TEM (mm) for selections of assessors.

	Assessors			
	1–2	1–3	2–3	1–2–3
$\mathrm{TEM}_{\mathrm{inter}}$	23.5	26.5	27.4	25.9

Table 3
Bias (mm) between different pairs of assessors.

	Assessors		
	1–2	1–3	2–3
Bias	+0.1	+22.2	+22.1

3.2.3. Intra-observer error

Assessor 2 has performed the point marking twice for all nine participants. This allows the intra-TEM to be measured as

$$\mathrm{TEM}_{\mathrm{intra}} = \left(\frac{1}{2n}\sum_{i=1}^{n}(s_i^1 - s_i^2)^2\right)^{1/2},$$

where s_i^1 and s_i^2 are the predicted statures obtained in the first and second round of marking, respectively. The intra-TEM may be interpreted as an indicator of the measurements' reliability. As measurements are repeated, some variations are initially expected until a point where the error is reduced to a small value and progress can no longer occur. When the intra-TEM stagnates, one can be confident about the predicted statures. For assessor 2, the $\mathrm{TEM}_{\mathrm{intra}}$ is found to be equal to 35.1 mm. This value suggests that possible improvement of the landmark positioning can be made.

3.2.4. Assessor's bias

The bias between two assessors placing markers can be quantified as

$$\mathrm{bias}_{a-b} = \frac{1}{n}\sum_{i=1}^{n}(s_i^a - s_i^b),$$

where s_i^a and s_i^b are the i-th predicted statures obtained from assessor a and b, respectively. The variables a and b take distinct values in the range $1,\ldots,k$. Results are summarised in Table 3. Again, the more experienced anatomists (pair 1–2) recorded much smaller bias than less experienced ones (pair 1–3). Looking at pair 1–2, the positive value for the bias means that on average the first assessor predicted taller statures than the second one. A similar reasoning can be deduced regarding the other two pairs.

3.2.5. Effects of garments

$\mathrm{TEM}_{\mathrm{truth}}$ provides an error measure which is too generic and does not reveal the effect of a particular garment on the assessors' ability to mark the required points. The analysis in this section addresses this limitation. First, the inter-TEM is calculated by including the measurements that only relate to a specific clothing style: (a) all participants are shirtless, (b) with a black shirt, (c) a striped shirt or (d) a padded jacket ($n = 9$). Table 4 presents the results. Overall, the TEM for shirtless participants turn out to be the largest due to the roundness of the shoulder and thus the increased ambiguity to mark the acromiale. Lowest TEM is achieved for participants wearing a black shirt as it defines the silhouette better around the shoulders. When looking at the various garment types, those with stripes or padded produce higher inaccuracies, which is to be expected.

Table 4
Inter-TEM (mm) for (a) shirtless participants; (b) with black shirt; (c) with striped shirt; (d) with padded jacket.

	Assessors			
	1–2	1–3	2–3	1–2–3
(a)	30.0	28.3	37.7	32.3
(b)	16.6	26.0	10.0	18.7
(c)	22.5	21.3	25.8	23.3
(d)	22.8	29.8	28.7	27.3

Table 5
Bias (mm) between different garment types. The abbreviation ‘SL’ stands for ‘Shirtless’.

	SL-black	SL-striped	SL-padded
Bias	+14.5	+36.4	+37.8

In a second series of tests, the bias is examined by comparing the shirtless case to the clothed ones. The formula in Section 3.2.4 is used with s_i^a taken as the stature measurement obtained for a shirtless participant and s_i^b as the measurement for the corresponding participant wearing either the black shirt, the striped shirt or the padded jacket for all assessors ($n = 27$). Results are given in Table 5.

As can be seen, the difference in marking anthropometric points is significantly smaller when shirtless participants are compared to those wearing a black shirt. The striped shirt and padded jacket increase the difficulty in identifying points which yields larger errors in both cases. The consistently positive bias across all three categories indicates that the estimated statures are taller on average for shirtless participants and thus may suggest a tendency to place markers more incorrectly in this situation. This would agree with the results in Table 4 where the inter-TEMs (almost) always show greater variation in the shirtless case than in the other three cases.

3.2.6. Snedecor’s F-test

Technical errors of measurements are essentially variances of one measurement around another measurement. Their random errors are thus a result of the measurement-to-measurement differences and sample sizes. This means that the difference between two TEM values can be tested for statistical significance in the same way the difference of two variances is. The Snedecor’s *F*-test is an appropriate tool to use. This test is based on the ratio of two variances in general populations to assess significance because the distribution of errors of ratios of larger to smaller variances depends on the combination of their degrees of freedom that are determined by sample sizes minus one. The *F*-test is given by

$$F = \frac{v_1}{v_2},$$

where v_1 is the larger variance, v_2 the smaller variance. Both v_1 and v_2 are estimates of population variances derived from sample values in the following way:

$$v_1 = \frac{v'_1 \cdot N}{N-1}, \quad v_2 = \frac{v'_2 \cdot N}{N-1}$$

with v'_1 and v'_2 the sample variances and N the number of observations. TEMs are square roots of sample variances, hence after squaring TEM values, multiplying them by N and dividing the result by $N - 1$, we obtain equivalents of variances appropriate to form ratios for the *F*-test. Note that with large numbers of observations, the $N/(N - 1)$ term approaches 1 and thus direct ratio of squared TEMs is an approximation of the *F*-test value.

For the TEMs discussed in Sections 3.2.1, 3.2.2 and 3.2.5, with the number of observations $N = 9$, most squared TEM ratios do not exceed appropriate cut-off *F*-test values for the 0.05 significance of differences. In Table 4, the TEMs in rows (a)–(b) for assessors 1–2 and 2–3 are the two instances where the *F*-tests are statistically different. This is because locating landmarks is much easier when a person wears a black shirt than no shirt, as explained in Section 3.2.5.

Table 6
TEM_{truth} (mm) when the dactylion and ground points are those from the expert and novice assessors.

	Assessor 1	Assessor 2	Assessor 3
TEM_{truth}^{exp}	30.0 (+0.0)	38.7 (−0.5)	32.9 (−11.4)
TEM_{truth}^{nov}	44.1 (+14.1)	48.9 (+9.7)	44.3 (+0.0)

3.3. Error analyses for a single marker

Among the three markers to place, only one of them is truly covered by clothing: the acromiale. In this section, we investigate the errors and effects of garments on this particular landmark. Since all three markers are recorded per assessor for all photographed subjects, we conducted a first series of tests whereby the dactylion and ground points are always taken as those from the *expert* assessor. This means that the location of the acromiale remains as chosen by the individual assessor. The error measures presented in Section 3.2 are labelled with a superscript ^{exp} to mark this distinction. We have also recalculated all the errors when the dactylion and ground points originate from the *novice* assessor. These are labelled with a superscript ^{nov}. The two data manipulation strategies are employed to examine how the errors fluctuate for radically different expertise levels and whether it reveals any pattern.

3.3.1. Comparison to truth

Table 6 summarises the results for TEM_{truth}. The values in brackets indicate the relative difference with the results in Table 1 when all three markers are chosen by each assessor.

Since Assessor 1 is the person with expert anatomical knowledge, his score for TEM_{truth}^{exp} remains unchanged from Table 1 (identical data). A considerable improvement can be noted for the novice anatomist (Assessor 3) who progresses to a comparable level to that of the expert assessor. The TEM for the trained anatomist decreased minimally. This overall trend is to be expected since Assessor 1 has a better selection of points (smallest error to truth in Table 1).

In the second row, the fixed dactylion and ground points come from Assessor 3 so his score is unchanged from Table 1. The increase in value for the results of the other assessors is understandable given that Assessor 3 has the largest discrepancy to truth when all three markers are specified.

Looking globally at the results, the relative difference between assessors 1 and 3 is negligible for both TEM_{truth}^{exp} and TEM_{truth}^{nov}. This suggests that they have consistently placed the acromiale around the same location. The difference in landmark placement is more noticeable for assessor 2 who has a larger residual error in both tests.

3.3.2. Inter-observer error

Table 7 shows the new inter-TEM values. The numbers in brackets indicate the difference with the results in Table 2 when all three markers are chosen freely by each assessor.

All errors have decreased and turned out about the same magnitude. The smaller variations between assessors are a direct consequence of fixing two of the three markers. Assessors 1–3 produced the smallest inter-TEM values whereas the largest values are observable when statures from assessor 2 are compared to

Table 7
Inter-TEM (mm) for selections of assessors.

	Assessors			
	1–2	1–3	2–3	1–2–3
TEM_{inter}^{exp}	18.6 (−4.9)	12.2 (−14.3)	17.5 (−9.9)	16.3 (−9.6)
TEM_{inter}^{nov}	18.7 (−4.8)	12.6 (−13.9)	17.1 (−10.3)	16.4 (−9.5)

those of assessors 1 and 3. Since the only source of uncertainty arises from the location of the acromiale, these results confirm that assessors 1 and 3 placed similar landmarks, and that assessor 2 was visually more affected by the clothing styles.

3.3.3. Assessors' bias

The bias is now calculated for the new data, refer to Table 8. All errors are small and about the same magnitude. Statures from assessor 3 are shorter than those of assessor 1 (positive bias) but taller than those of assessor 2 (negative bias) with similar amount of variation in each case. From column 1, assessor 2 yielded shortest statures. This agrees with prior findings that this assessor has tangibly different point locations than the other assessors.

Assuming equal difficulty in marking the acromiale and dactylion, one may deduce from the bias values in Tables 3 and 8 that a large part of the error stems from the location of the ground point. Assessor 2 must have placed this point much better than assessor 3 because all relevant errors in Section 3.2 are larger for assessor 3 and we have identified in this section that his placement of the acromiale is comparable to the expert anatomist.

Table 8
Bias (mm) between different pairs of assessors.

	Assessors		
	1–2	1–3	2–3
Bias^{exp}	+3.8	+2.0	−1.8
Bias^{nov}	+4.0	+1.8	−2.2

3.3.4. Effects of garments

Following the analysis in Section 3.2.5, we consider the inter-TEMs for statures that only relate to a particular type of clothing. Results are summarised in Tables 9 and 10.

Performing a row-wise comparison between the two tables, we see that all errors have about the same order of magnitude and have decreased compared to their corresponding values in Table 4. In relation to clothing effects, the decrease is more significant for striped shirt and padded jacket which demonstrates that these garments present a greater challenge for someone to place markers correctly. This corroborates the conclusion in Section 3.2.5.

Other experiments focus specifically on the errors between different clothing styles. Looking at Table 11, the effect of a particular garment is clearly visible since most often the bias is positive with large magnitude. The trends are shown more prominently here compared to Table 5 because the uncertainty is assessed precisely in the marker covered by clothing. Note that the bias in column 1, row 2, is small and negative. This peculiarity is a consequence of using the ground points from assessor 3, which we know from previous experiments are not well placed. This result may be ignored.

Table 9
$\mathrm{TEM}^{exp}_{inter}$ (mm) for (a) shirtless participants; (b) with black shirt; (c) with striped shirt; (d) with padded jacket.

	Assessors			
	1–2	1–3	2–3	1–2–3
(a)	26.6	14.2	16.2	19.8
(b)	13.4	15.8	13.3	14.2
(c)	13.0	8.7	18.5	14.0
(d)	17.9	8.3	21.0	16.7

Table 10
$\mathrm{TEM}^{nov}_{inter}$ (mm) for (a) shirtless participants; (b) with black shirt; (c) with striped shirt; (d) with padded jacket.

	Assessors			
	1–2	1–3	2–3	1–2–3
(a)	27.0	14.2	16.0	19.9
(b)	14.1	16.5	13.6	14.8
(c)	12.2	8.7	17.3	13.2
(d)	17.9	9.1	20.9	16.7

Table 11
Bias (mm) between different garment types. The abbreviation 'SL' stands for 'Shirtless'.

	SL-black	SL-striped	SL-padded
Bias^{exp}	+33.4	+45.1	+34.7
Bias^{nov}	−5.3	+31.8	+38.0

3.3.5. Snedecor's F*-test*

F-tests have been performed in a similar fashion to those in Section 3.2.6. Looking at Tables 2 and 7, the TEMs for the assessors' pair 1-3 exhibit statistically significant differences at the 95% confidence level. This result is a formal confirmation that the provision of better ground points has improved the stature estimates of assessor 3 and thus reduced the inter-TEM with assessor 1 noticeably.

Other statistically significant differences can be found in both Tables 9 and 10 between rows (a)–(b) and (a)–(c) for assessors' pair 1–2, between rows (b)–(d) for pair 1–3 in Table 9 and rows (b)–(c) for pair 1–3 in Table 10. Since acromiale is the only variable point, the *F*-tests prove that assessors 1–2 have consistently placed that point when participants wear shirts (low inter-TEMs in rows (b), (c)) compared to no shirt (row (a)). For assessors 1–3, the *F*-tests reveal that they had similar point marking for participants wearing a striped shirt and a padded jacket (rows (c), (d)) compared to when they wear a black shirt (row (b)).

3.4. Real-life surveillance images

Our model is now applied to an airport surveillance video released for the Performance Evaluation of Tracking and Surveillance 2007 workshop [38]. We have examined 4500 images (of resolution 720 × 576 pixels) of a particular video clip. Over 95% of passengers and bypassers are found to be missing lower limbs due to luggages and other people obstructing. To have benchmark statures for comparison, we have considered the same 30 men as those in [13], however several challenges appeared immediately:

1. Some bypassers wear clothes of colour similar to the scene background. This especially precludes the marking of the acromiale;
2. Many waiting passengers stand with their arms crossed over their chest or behind their back, their upper limb is partially occluded, their hands are closed or hidden;
3. Image resolution is too poor to clearly distinguish body parts of pedestrians far in the scene. Markers would need to be placed with sub-pixel accuracy, which is not straightforward to do.

In these situations, the point placement is not trustworthy or possible, so the men are discarded. Such situations are illustrated in Fig. 4. This brings the number of test subjects down to 18. Two of these subjects are viewed facing the camera, one from the back and fifteen others under various side-way postures. The latter postures are most difficult to deal with, even when the complete upper limb is visible (Fig. 5). It is indeed easier to locate the acromiale when both shoulders are observable as in our laboratory experiments where participants are in anterior or posterior view. Among the

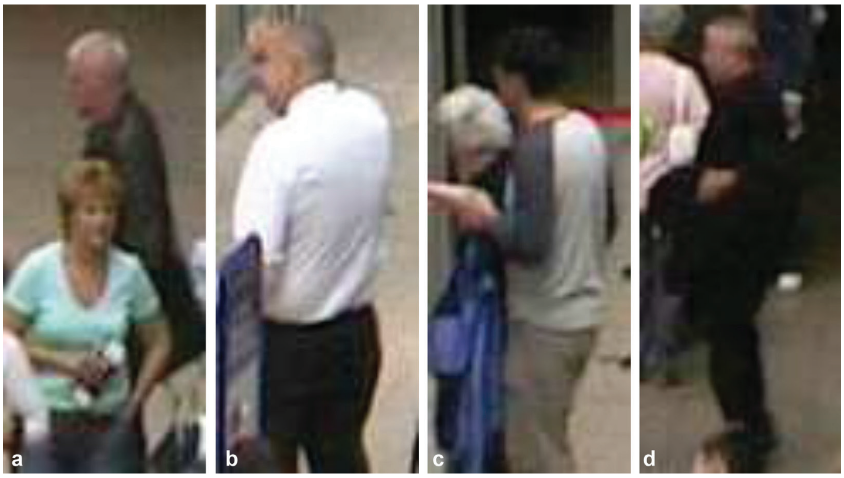


Fig. 4. Examples of subjects discarded from our experiments.

test subjects, four passengers have garments of colour similar to the background and two others walk with their arms bent pushing trolleys. We kept these six candidates to see how the assessors and our model would cope with extreme situations. Figs. 5 and 6 depict some workable examples and other more challenging cases in our test set.

In order to carefully examine the effects of garments, the 18 selected men are separated into six categories of equal size. In this context, garment style but also subject posture must be taken into account, the latter being a novel addition compared to the laboratory experiments. People for whom the upper limb is reasonably visible are considered as having an adequate posture; otherwise, they are labelled as having inadequate posture. The categories are:

(a) *Easy-shirt*: people are wearing a shirt and stand with adequate posture;
(b) *Easy-jumper*: people are wearing a jumper and stand with adequate posture;
(c) *Easy-jacket*: people are wearing a padded jacket or thick coat and stand with adequate posture;
(d) *Hard-shirt*: people are wearing a shirt and stand with inadequate posture;
(e) *Hard-jumper*: people are wearing a jumper and stand with inadequate posture;
(f) *Hard-jacket*: people are wearing a padded jacket or thick coat and stand with inadequate posture;

3.4.1. *Experimental arrangements*

In the test images, all 18 passengers have partially occluded legs and feet, so no ground point is markable. The authors in [13] have kindly provided the camera calibration, a list of subjects, their predicted statures from head heights, top and base head point

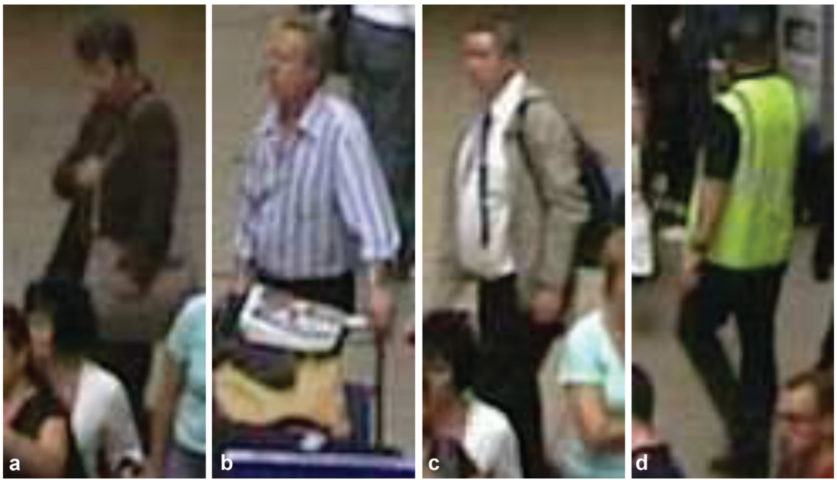


Fig. 5. Subjects with visible upper limb.

Fig. 6. Challenging situations for point marking.

locations and their projections onto a reference plane. Our experimental set-up could thus replicate their exact conditions. We show next how the ground point entering the calculation of the upper limb length is obtained from this starting information.

The world coordinate system is set such that the *X*–*Y* plane is on the ground and the positive *Z*-axis represents the upward vertical scene direction. Let $\mathbf{o}_w$ be the image of the world origin, $\mathbf{t}$ the head top point and $\tilde{\mathbf{t}}$ its projection onto a reference plane perpendicular to the ground plane (Fig. 7). Furthermore, let $\mathbf{v}_x$, $\mathbf{v}_y$, $\mathbf{v}_z$ denote the vanishing points in the *X*-, *Y*-, *Z*-directions, respectively. Viewing the 3-D world as a collection of three orthogonal pencils of parallel planes [22], it can be shown that the projection of $\tilde{\mathbf{t}}$ onto the ground plane, which is aligned with $\mathbf{t}$ and $\mathbf{v}_z$, is the homogeneous point

$$\mathbf{g} = ([\mathbf{t}^T, 1]^T \times \nu_z) \times (\nu_x \times \mathbf{m})$$

with $\mathbf{m} = ([\mathbf{o}_w^T, 1]^T \times \nu_y) \times ([\tilde{\mathbf{t}}^T, 1]^T \times \nu_z)$. Point $\mathbf{g}$ is then projected orthogonally onto the guide line formed by the (ML) acromiale and dactylion points (Fig. 7). This technique yields a valid ground point for upper limb measurement. Only two points are now required, the acromiale and the dactylion. As in the laboratory experiments, assessors have marked each point in a single action. This is a major improvement over methods which require extensive repeats of the point placement and need to operate at sub-pixel level [13].

3.4.2. Test results

The same three assessors as those in Section 3.1 have placed markers to obtain the upper limb lengths for all subjects. These lengths are subsequently used to infer statures through OLS and RMA regressions (Section 2.4). The final statures are taken as the average values of the two predictions to match the approach in [13] and compare estimates. Figs. 8 and 9(a) show the results for all three assessors along with the predicted statures from [13].

One may consider the statures from [13] as “truth” and obtain a technical error of measurement using the assessors’s body heights and the formula given in Section 3.2.1. As can be seen from Table 12, assessor 3 produced the lowest score. This surprising

Fig. 7. Construct for upper limb length measurement with point $\tilde{\mathbf{t}}$ (magenta) on the reference plane (green), $\mathbf{t}$ (cyan) on the head vertex, $\mathbf{g}$ (yellow) on the ground and its projection (red) onto the guide line.

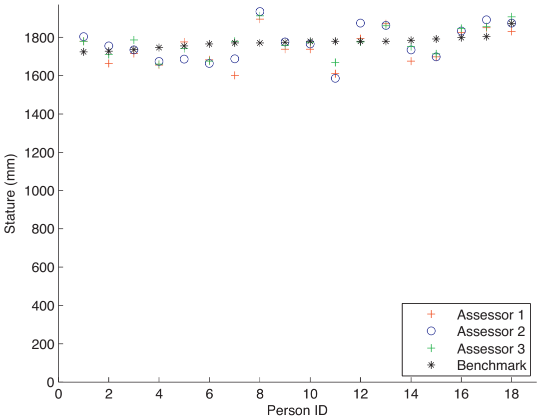
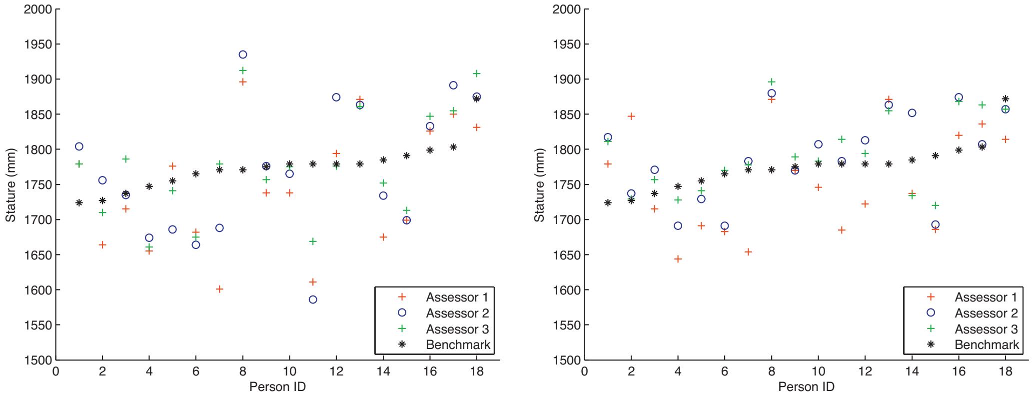


Fig. 8. Stature estimates for 18 passengers. Benchmark statures originate from [13].



(a) Test results (b) Re-test results

Fig. 9. Stature estimates shown on a macro scale. Benchmark statures originate from [13].

Table 12
Errors (mm) between predicted statures and [13].

	Assessor 1	Assessor 2	Assessor 3
$\text{TEM}_{\text{truth}}$	60.8	61.2	45.3

result may be explained by the fact that assessor 3 gained familiarity with the subjects as a consequence of spending a significant amount of time setting up the experimental images. We also believe that the laboratory experiments have been beneficial for training. His landmarks are therefore better placed, in particular for subjects standing with their upper limb at an angle from their body (e.g. when pushing a trolley). The other two assessors performed equally to each other, although with larger errors than assessor 3.

The assessors' inter-TEM and bias have been tested as per Sections 3.2.2 and 3.2.4, refer to Table 13. Given that assessors 2 and 3 have obtained better scores for $\text{TEM}_{\text{truth}}$, their inter-TEM and bias values are lower than those of other combinations of assessors. The negative bias values suggest that assessor 1 has obtained shorter statures on average compared to the other assessors.

3.4.3. *Re-test results*

A re-test session was organised to calculate the intra-observer TEM and observe the variations of other errors. So, the three assessors have repeated the point marking a second time for all subjects. Results are shown in Tables 14 and 15 with an accompanying graph in Fig. 9(b).

Compared to TEMs in Table 12, the new values show improvement in landmark localisation and therefore stature estimation. Two of the assessors have scored TEMs equivalent to those in the laboratory experiments, which is encouraging. The average stature from [13] is 1775 mm. Considering the worst error (assessor 1), this still gives an accuracy of about 97%. Looking at the intra-TEMs reveals that assessors have either become more precise in their marking or changed their approach (indicated by the large values). This is supported by the variations of predicted statures in the graphs of Fig. 9. These intra-TEMs also suggest that assessors have a margin of progress.

According to Table 14, assessors 2 and 3 have close stature estimates (from $\text{TEM}_{\text{truth}}$), which implies that their landmark positions may be similar. In turn, this means they should produce a lower inter-TEM and bias compared to other pairs of assessors. This is indeed confirmed in Table 15. Results of assessor 1 are about as far apart from each of the other two assessors. The negative bias confirms that assessor 1 generally produces shorter statures as can be seen in Fig. 9(b). Overall, the two graphs of Fig. 9 show that assessors' predicted statures after re-test are less spread out, which is expected as they repeat the experiments.

Table 13
Inter-TEM and bias (mm) for selections of assessors.

	Assessors			
	1–2	1–3	2–3	1–2–3
$\text{TEM}_{\text{inter}}$	34.9	40.3	32.3	36.0
Bias	−24.3	−30.8	−6.5	

3.4.4. *Effects of garments*

In the previous sections, $\text{TEM}_{\text{truth}}$ gives some global value with no distinction about the garment type. Using the statures obtained from the re-test experiments, we follow the same analytical process as in Section 3.2.5. The distribution of $\text{TEM}_{\text{truth}}$ per garment style and posture is summarised in Table 16. "Truth" is again taken as the predicted statures in [13]. The results for assessor 1 show that the error values are generally increasing with the complexity in garment style and subject posture. An inconsistency exists in row (d) where the error is abnormally large. This is rationalised by the fact that two of the three subjects in this category are pushing a trolley which creates an ambiguous situation, see Fig. 6(a). The dactylion can be marked near the hand on the trolley or approximately half way down the thigh (according to the upper limb model of Fig. 1). Assessor 1 decided on the former approach. The resulting upper limb lengths turn out much shorter than their

Table 14
Quantification of errors (mm) after re-test.

	Assessor 1	Assessor 2	Assessor 3
$\text{TEM}_{\text{truth}}$	53.5	40.9	36.5
$\text{TEM}_{\text{intra}}$	41.0	48.3	34.1

Table 15
Inter-TEM and bias (mm) after re-test.

	Assessors			
	1–2	1–3	2–3	1–2–3
TEM_{inter}	46.4	47.1	28.1	41.5
Bias	−35.9	−39.8	−3.9	

actual lengths due to the bent elbow, hence the large TEM value. Assessors 2 and 3 opted for the latter approach. They have obtained greater lengths and consequently lower errors for this category and in row (e). Our current model would benefit from a multiple-part regression for resolving the ambiguity. Although this is not the intended focus of the present work, future extensions to accommodate the issue are possible and discussed in Section 4.

Aside from these results, we also observe increasing error values (or very similar values) in other categories of assessors 2 and 3. Drawing special attention to row (e) of assessor 3, the error turned out very small. Investigation revealed that some of the images here are those that were used for developing the model. So, assessor 3 subconsciously gained familiarity with the subjects. Some level of progress is also expected from the first round of marking. This result is very powerful in that it indicates the extent to which the error may be decreased.

Table 17 presents the inter-TEMs for measurements which relate to specific garment types and subject postures. Most variations (largest errors) occur in the three hard-cases categories (rows d–f), especially in row (d) for the pairs 1–2 and 1–3. This is expected since assessor 1 has placed the landmarks most differently from the other two assessors for subjects in these classes. The inter-TEM for the pair 2–3 in row (d) is understandably smaller since these assessors have followed the same marking strategy. Performing Snedecor *F*-tests with $N = 18$ between rows (a)–(d), (a)–(e) and (a)–(f) of pairs 1–2 and 1–3 confirm that there are statistically significant differences at the 0.05 significance level.

Overall, although results have improved from the first experiments, substantial differences are present. Indeed, even when considering the best results (for assessors 2–3), squared inter-TEM ratios exceed cut-off *F*-test values. This proves that there still exists significant variations in the measurements and therefore all errors should be reducible further. This analysis agrees with the conclusion from examining the intra-TEMs (Table 14).

The bias between garments (which includes subject posture) is calculated by selecting statures from the first category (easy-shirt) and comparing to other categories for all assessors, see Table 18. Clearly, the effects on stature estimation become increasingly important as the garment style gets thicker and the subject posture is more complicated. The inconsistency for the pair (a)–(d) simply reflects the incorrect measurements of assessor 1 as seen in row (d), column 1 of Table 16.

Table 16
TEM_{truth} (mm) per assessor based on the garment style and subject posture. Categories (a) to (f) are described at the start of Section 3.4.

	Assessor 1	Assessor 2	Assessor 3
(a)	27.3	13.1	8.5
(b)	37.0	39.5	43.5
(c)	44.6	52.3	41.2
(d)	74.3	23.4	16.5
(e)	63.7	33.5	6.5
(f)	58.8	62.5	63.2

Table 17
Inter-TEM (mm) for groups of assessors based on the garment style and subject posture.

	Assessors			
	1–2	1–3	2–3	1–2–3
(a)	30.5	24.4	12.5	23.7
(b)	24.9	26.6	23.5	25.1
(c)	31.9	29.5	12.7	26.1
(d)	68.9	80.7	19.8	62.3
(e)	58.4	66.7	33.3	54.7
(f)	47.2	12.2	48.7	39.8

Table 18
Bias (mm) between different garment style and subject posture.

	(a)–(b)	(a)–(c)	(a)–(d)	(a)–(e)	(a)–(f)
Bias	+13.2	+32.1	+70.3	+45.2	−40.7

4. Discussion

The experiments have shown that, despite landmark positioning errors, the accuracy of height estimates from real-life CCTV images can be commensurate with, if not surpass, the expected 95% accuracy of height reconstruction from direct measurement of skeletal remains. This has been a long challenge. The use of the upper limb length compared to the head height [13] has decreased user interaction to a single action or two instead of extensive repeats to guarantee equivalent accuracy. This claim is supported by the results of assessors coming from a range of backgrounds in anatomy and computer vision. Variations between assessors suggest that some training in photoanthropometry is beneficial to reduce marking errors—*Practice makes perfect*.

We infer from the results that at present the weakest component in our procedure lies in the regression model, not the human factor. Large international databases of body measurements exist through the Civilian American and European Surface Anthropometry Resource (CAESAR) project [39]. We intend to use these anthropomeasures to construct improved regression equations. This would include multiple regressions from various body dimensions to refine our current upper limb model and combine it with other ones such as [13] to allow for a more complete characterisation of human beings.

This research has permitted us to learn about the extent of errors involved in precise body part measurement from real-life surveillance imagery. The lessons may now be applied to enhance body-part acquisition from an automated detector or tracker. In addition, this knowledge could help in improving the recognition rate when matching subjects in uncontrolled scenes to ideal CAESAR data [19].

5. Conclusion

This paper has examined the effects of various types of garments on human stature estimation from images. To this end, we have developed a procedure whereby the upper limb length is first measured from the image and then stature is inferred by linear regression from it. Three assessors have experimentally marked upper limb points of subjects in both laboratory and real-life surveillance videos. In both scenarios, thicker garments and those with stripes produce higher inaccuracies in stature prediction, which is to be expected. Seven error measures are used to study the variations between obtained statures. The most valuable outcome is that errors are within the expected variance of the stature regressor. Thus, body heights from imaged upper limbs can be inferred with confidence that is no worse than the accuracy of reconstructed body parts in routine skeletal forensic work.

References

[1] M. Yoshino, K. Noguchi, M. Atsuchi, S. Kubota, K. Imaizumi, C. David, L. Thomas, J.G. Clement, Individual identification of disguised faces by morphometrical matching, Forensic Sci. Int. 127 (2002) 97–103.

[2] A. Burton, S. Wilson, M. Cowan, V. Bruce, Face recognition in poor-quality video: evidence from security surveillance, Psychol. Sci. 10 (1999) 243–248.

[3] S. Chen, M.H.S. Mau, C. Sanderson, A. Bigdeli, B. Lovell, Face recognition from still images to video sequences: a local-feature-based framework, Int. J. Image Video Process. (2011) 1–14.

[4] P. Kovesi, Video surveillance: legally blind? in: Proceedings of International Conference on Digital Image Computing Techniques and Applications, 2009, pp. 204–211.

[5] M. Henneberg, Facial mapping, body mapping and the duties of an expert witness, 2007, http://www.lawlink.nsw.gov.au/lawlink/pdo/ll_pdo.nsf/pages/PDO_facial-mapping.

[6] M. Henneberg, Expert witness in a courtroom: the Australian experience, in: M. Oxenham (Ed.), Forensic Approaches to Death Disaster and Abuse, Australian Academic Press, Queensland, 2008.

[7] A. Bertillon, Identification anthropométrique. Méthode nouvelle de détermination de l'identité individuelle, Masson, Paris, 1883.

[8] A. Bertillon, La photographie judiciaire, avec un appendice sur la classification et l'identification anthropométriques, Gauthier-Villars, Paris, 1890.

[9] C. Barron, I.A. Kakadiaris, Estimating anthropometry and pose from a single uncalibrated image, Comput. Vis. Image Underst. 81 (3) (2001) 269–284.

[10] C. BenAbdelkader, L.S. Davis, Estimation of anthropomeasures from a single calibrated camera, in: Proceedings of International Conference on Automatic Face and Gesture Recognition, 2006, pp. 499–504.

[11] C. BenAbdelkader, Y. Yacoob, Statistical body height estimation from a single image, in: Proceedings of International Conference on Automatic Face and Gesture Recognition, 2008, pp. 1–7.

[12] D. Cao, C. Chen, M. Piccirilli, D. Adjeroh, T. Bourlai, A. Ross, Can facial metrology predict gender? in: Proceedings of International Joint Conference on Biometrics, 2011, pp. 1–8. , http://dx.doi.org/10.1109/IJCB.2011.6117471.

[13] T. Scoleri, M. Henneberg, View-independent prediction of body dimensions in crowded environments, in: Proceedings of International Conference on Digital Image Computing Techniques and Applications, 2012, pp. 1–8.

[14] N. Rudd, V. Chattaraman, Preferences for aesthetic attributes in clothing as a function of body image, body cathexis and body size, Cloth. Textiles Res. J. 24 (1) (2006) 46–61.

[15] H. Firth, K. Gleeson, Clothing and embodiment men managing body image and appearance, Psychol. Men. Masc. 4 (2004) 40–48.

[16] M. Johnson, Appearance modifying garments, U.S. Patent 5,065,458, 1991.

[17] P. Thompson, K. Mikellidou, Applying the Helmholtz illusion to fashion: horizontal stripes won't make you look fatter, i-Perception 2 (1) (2011) 69–76.

[18] J. Fan, W. Yu, L. Hunter, Clothing Appearance and Fit: Science and Technology, Woodhead Publishing, Cambridge, 2004.

[19] R. McCoppin, M. Rizki, L. Tamburino, A. Freeman, O. Mendoza-Schrock, The effects of clothing on gender classification using LIDAR data, in: IEEE National Aerospace and Electronics Conference (NAECON), 2012, 134–139. , http://dx.doi.org/10.1109/NAECON.2012.6531043.

[20] T. Lucas, The extent to which garments affect the assessment of somatotypes of adult South Australian males from CCTV images, School of Medical Sciences, University of Adelaide, 2012 November (Honours thesis).

[21] D. De Angelis, R. Sala, A. Cantatore, P. Poppa, M. Dufour, M. Grandi, C. Cattaneo, New method for height estimation of subjects represented in photograms taken from video surveillance systems, Int. J. Legal Med. 121 (6) (2007) 489–492.

[22] A. Criminisi, Accurate visual metrology from single and multiple uncalibrated images, University of Oxford, Dept. Engineering Science, 1999 December (Ph.D. thesis), http://www.robots.ox.ac.uk/~vgg.

[23] G. Edelman, I. Alberink, B. Hoogeboom, Comparison of the performance of two methods for height estimation, J. Forensic Sci. 55 (2) (2010) 358–365.

[24] N.H. Nguyen, R. Hartley, Height measurement for humans in motion using a camera: A comparison of different methods, in: Proceedings of International Conference on Digital Image Computing Techniques and Applications, 2012, pp. 1–8.

[25] J.K. Aggarwal, Q. Cai, Human motion analysis: a review, Comput. Vis. Image Underst. 73 (1999) 428–440.

[26] C. BenAbdelkader, R. Cutler, L.S. Davis, Person identification using automatic height and stride estimation, in: Proceedings of International Conference on Pattern Recognition, vol. 4, 2002, pp. 377–380.

[27] R.T. Collins, R. Gross, J. Shi, Silhouette-based human identification from body shape and gait, in: Proceedings of International Conference on Automatic Face and Gesture Recognition, 2002, pp. 366–371.

[28] S.-H. Lee, J.-S. Choi, A single-view based framework for robust estimation of height and position of moving people, in: Proceedings of Pacific Rim Conference on Advances in Image and Video Technology, 2007, pp. 562–574.

[29] S. Blau, D.H. Ubelaker, Handbook of Forensic Anthropology and Archaeology, Left Coast Press, Walnut Creek, CA, 2009.

[30] S.M. Byers, S.M.T. Myster, Forensic Anthropology, Pearson, Boston, 2005.

[31] P. Willey, Stature estimation, in: S. Blau, D.H. Ubelaker (Eds.), Handbook of Forensic Anthropology and Archaeology, Left Coast Press, Walnut Creek, CA, 2009.

[32] K. Krishan, K. Vij, Diurnal variation of stature in three adults and one child, Anthropologist 9 (2) (2007) 113–117.

[33] L.D. Voss, B.J.R. Bailey, Diurnal variation in stature: is stretching the answer? Arch. Dis. Child. 77 (1997) 319–322.

[34] M.J. Brooks, W. Chojnacki, D. Gawley, A.V.D. Hengel, What value covariance information in estimating vision parameters, in: Proceedings of International Conference on Computer Vision, IEEE Computer Society, 2001, pp. 302–308.

[35] R. Martin, K. Saller, Lehrbuch der Anthropologie, Gustav Fischer Verlag, Stuttgart, 1957.

[36] T. Scoleri, Video metrology without the image-to-ground homography, in: Proceedings of International Conference on Digital Image Computing Techniques and Applications, 2010, pp. 335–342.

[37] S.J. Ulijaszek, J.A. Lourie, Intra- and inter-observer error on anthropometric measurement, in: S.J. Ulijaszek, C.G.N. Mascie-Taylor (Eds.), Anthropometry: The Individual and the Population, Cambridge University Press, Cambridge, 1994.

[38] PETS, Performance Evaluation of Tracking and Surveillance Benchmark Data, 2007, http://pets2007.net/.

[39] Civilian American and European Surface Anthropometry Resource (CAESAR) database, http://store.sae.org/caesar/.

Appendix 2: Reprint of ‘Are human faces unique? A metric approach to finding single individuals without duplicates in large samples’.

Forensic Science International 257 (2015) 514.e1–514.e6

Contents lists available at ScienceDirect

Forensic Science International

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

Forensic Anthropology Population Data

Are human faces unique? A metric approach to finding single individuals without duplicates in large samples

Teghan Lucas *, Maciej Henneberg

Biological and Comparative Anatomy Research Unit, Medical School, University of Adelaide, Adelaide 5005, Australia

ARTICLE INFO

Article history:
Received 21 November 2014
Received in revised form 21 May 2015
Accepted 7 September 2015
Available online 25 September 2015

Keywords:
Forensics
Human variation
Identification
Singularity
Forensic anthropology population data

ABSTRACT

In the forensic sciences it is inferred that human individuals are unique and thus can be reliably identified. The concept of individual uniqueness is claimed to be unprovable because another individual of same characteristics may exist if population size were infinite. It is proposed to replace “unique” with “singular” defined as a situation when only one individual in a specific population has a particular set of characteristics. The likelihood that in a population there will be no duplicate individual with exactly the same set of characteristics can be calculated from datasets of relevant characteristics.

To explore singularity, the ANSUR database which contains anthropometric measurements of 3982 individuals was used. Eight facial metric traits were used to search for duplicates. With the addition of each trait, the chances of finding a duplicate were reduced until singularity was achieved. Singularity was consistently achieved at a combination of the maximum of seven traits. The larger the traits in dimension, the faster singularity was achieved. By exploring how singularity is achieved in subsamples of 200, 500, etc. it has been determined that about one trait needs to be added when the size of the target population increases by 1000 individuals. With the combination of four facial dimensions, it is possible to achieve a probability of finding a duplicate of the order of 10^{-7}, while, the combination of 8 traits reduces probability to the order of 10^{-14}, that is less than one in a trillion.

1. Introduction

In the forensic sciences the process of identification relies upon a finding that a trace left at the crime scene (or other location relevant to the investigation) and the suspected source, often an object or a person, correspond to each other in essential characteristics. In brief, they match. In order to make a match between a trace and a source that allows conclusion of identification, the probability of finding a duplicate trace or a duplicate source must be negligible. If this probability equals zero the match is unique [1]. However, it can be argued that unless the whole world is included in searches for duplicates, the probability cannot be firmly established as zero [2]. This makes the term “unique” debatable. In real situations populations of traces and of sources are not infinite, thus probabilities of finding a duplicate can only approach zero. In those cases, when the probability of finding a duplicate is less than what the population size predicts, the actual match between the trace and the source is a single occurrence [1]. We propose to call this match a “singularity”. In everyday terms it could be called a “unique correspondence of trace and source essential characteristics”, but the ambiguity of the term “uniqueness” remains a problem. We define singularity as the correspondence between essential characteristics of the trace and the source that in a given population has a probability of occurrence less than that predicted from random combination of characteristics of the trace and the source. This probability predicts that no duplicate of the trace or of the source can be found in the given population.

The term “singular” as defined here is free from the ambiguity of the word “unique” and thus may be more appropriate to use in forensic statements and court proceedings. Unlike statements saying that a particular individual is unique, the statement that the individual is singular in a defined population is easily testable both empirically and in court proceedings.

The lower the probability of finding a duplicate trace the more reliable the evidence is considered. Some widely recognised claims of unique traces left by humans are: DNA [3], fingerprints [4,5], bite marks [6]. The lesser known traces of this kind include: elbow prints [7], ear prints [8], lip prints [9] and behavioural characteristics such as gait [10] and handwriting [11]. Uniqueness of these

---

* Corresponding author. Tel.: +61 8 8313 6324.
E-mail addresses: teghan.lucas@adelaide.edu.au (T. Lucas), maciej.henneberg@adelaide.edu.au (M. Henneberg).

http://dx.doi.org/10.1016/j.forsciint.2015.09.003
0379-0738/

traces is claimed on the principle that only one individual would be a source of such trace. This is to be criticised because, theoretically, another individual could be born with the same essential characteristics if we wait long enough. In practice, what is considered uniqueness of such traces is actually a singularity since populations from which those traces emanate are limited by time and space and thus are of a finite size.

The word 'traces' can refer to light rays producing changes on photographic film or on light sensitive digital photograph chips, that is to images. In recent years photographic traces have become increasingly popular as forensic evidence. In cases of morphological analyses, an expert witness specialising in the field of biological anthropology is often called upon to provide evidence of a match or mismatch between an image of an individual (a trace) and a suspect (a source) based on anatomical similarities of morphological traits. The current method for analysing image based traces is to use categorical scales of morphological traits [12–14]. For example, body height may be described as short, medium or tall. This method has been highly criticised on the grounds that it is not accurate and reliable [15–17]. Attempts have been made to address criticisms by attempting to take measurements from images, however, this research is still in progress [18,19].

Over the last few decades, facial recognition systems have been increasingly researched due to the proliferating use of images for identification. Some studies explain that the face is used as humans are good at recognising facial features [20,21], others say that obtaining images of the face is cheap and non-invasive [22]. According to Jafri and Arabnia [22] facial recognition is used for two primary tasks, verification (one to one matching) and identification (one to many matching).

There are a number of different scenarios where face recognition can be used including: passports, drivers licences, security, surveillance, etc. Some factors that make identification from images difficult include: illumination, facial expression, pose, distortions and pixelation [23]. Therefore research has concentrated on eliminating the confounding effects from images when making an identification.

There are two ways to analyse facial traits, descriptives and metrics. The use of descriptive traits involves adjectives such as 'wide' and 'curved' to categorise facial features such as the nose. Metric traits most commonly involve measuring the distances between specific points on the face. Theoretically, any descriptive trait, such as, for instance, face shape, can be converted to a metric one by taking measurements of its constituent properties such like the width, the height, the curvature, angles between its parts, etc. As mentioned earlier, descriptives are not considered as a reliable method of evaluation, especially in court proceedings [15–17]. Therefore, many facial recognition systems and methods have concentrated on moving towards metrics [24] or a combination of metrics and descriptives [25,26]. Facial recognition methods which use metrics alone or metrics and descriptives report high accuracy rates (>95%) [27,45]. However, each method of finding a match between a photograph and a person is tested on a small sample of subjects i.e. less than one hundred. Many photographs are used and a large number of measurements taken, but the number of people appearing on these photographs is limited. Small sample size automatically assures that no duplicate matches are found within the sample. These studies are based on the assumption that a face is unique, but very few justify this assumption quantitatively. In those studies that mention 'uniqueness' of the face, the term is not referenced nor defined. Therefore it seems that many people believe that the face is adequate to be used as a biometric tool, without it being sufficiently studied, especially when using metrics. A number of papers have identified the lack of knowledge in this area [28,29,25].

The aim of this study is to investigate whether or not two or more faces within a specified population have the same combination of several measurements. A secondary aim is to calculate the probability of not finding more than one face with same measurements (a duplicate) using a defined number of measurements in order to find the minimum number of facial dimensions needed to achieve singularity. We are not aiming here to investigate the accuracy with which measurements can be taken of various images and how such accuracy may influence findings of singularity. We only aim to introduce the principle of finding singularity that can be applied to any traces or sources, while the precision of its application requires a separate discussion of measuring techniques that may differ from case to case.

2. Materials and method

The U.S. Army Anthropometric Survey (ANSUR) database is a result of an anthropometric survey conducted in 1988 of U.S. military personnel. The dataset contains 132 manually measured dimensions of the human body and head. The sample consisted of 1774 men and 2208 women aged 17–51 years. This dataset was chosen as it is a sample of anthropometric measurements covering a range of variation of facial features. Even though the survey was conducted in 1988, it is still valid for the purposes of this study because the human face has not varied statistically within that time. Details of this study are described in ANSUR [30,44]. Initially males and females were analysed separately to see if sex influenced the numbers of metric traits needed to have no duplicates. Males and females were then combined to increase sample size and because sexual dimorphism accounts for only 25% variation in major measurable characters [31]. No further separation of the dataset (i.e. population of origin) was included as upwards of 95% of variation occurs between two randomly selected individuals rather than between individuals of different populations [32].

A team of 22 individuals conducted all measurements on the sample. The team was trained in anthropometry over a four week period. During this time, each member of the team was allocated specific measurements to learn and repeat continuously. These dimensions were then measured on the 3982 participants during data collection. By allocating specific measurements to each measurer, the measuring team aimed to reduce measurement errors. Measurement errors were calculated and reported alongside the database. Measurement errors ranged between 2.2% and 2.4% which is very small. For the purposes of this paper, measurement errors will not affect the results.

The following face/head measurements were used (Fig. 1): *Bitragion submandibular arc* – The surface distance between the right and left tragion across the submandibular landmark at the juncture of the jaw and the neck was measured with a measuring tape. *Bizygomatic breadth* – The maximum horizontal breadth of the face between the zygomatic arches was measured with a spreading caliper. *Ear length* – The length of the right ear is measured with a sliding caliper from its highest to lowest points on a line parallel to the long axis of the ear. *Ear protrusion* – The horizontal distance between the mastoid process and the outside edge of the right ear at its most lateral point (ear point) is measured with a sliding caliper. *Head breadth* – The maximum horizontal breadth of the head above the ears is measured with a spreading caliper. *Head circumference* – a measuring tape is used to measure the maximum circumference of the head above the supraorbital ridges and ears. *Head length* – the distance from the glabella landmark between the brow ridges to opisthocranion is measured with a spreading caliper. *Interpupillary distance* – a pupillometer is used to measure the distance between the centres of the right and

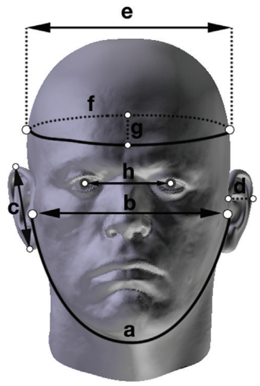


Fig. 1. Anterior view of the head/face showing: (a) bitragion submandibular arc; (b) bizygomatic breadth; (c) ear length; (d) ear protrusion; (e) head breadth; (f) head circumference; (g) head length; (h) interpupillary distance.

left pupils. The dimensions are illustrated in Fig. 1. The dimensions that are not influenced by facial expression were chosen.

Duplication is defined here as '*a situation where another trace is found matching a given trace*' since we have at our disposal only measurements (traces) of participants in the survey. In real forensic situations dimensions of a trace can be compared with dimensions of the actual person – the source. Depending on the number of traits measured on a trace and the accuracy with which they are measured, the ease of finding a duplicate will vary. If only a combination of two metric traits is used, it will be easier to find duplicates than when more dimensions will be used to characterise traces. As long as dimensions are not perfectly correlated with others ($r < 1.00$), adding a dimension should reduce the number of possible duplicates. In a defined sample, when a larger number of combinations of metric traits (traces) is used, a situation can be reached where no duplicates will be found. In the case of identification, a trace left behind by a particular individual can be analysed and described by a number of metric traits. This trace can be compared to the traces produced by the same measuring methods of any number of possible suspects from a defined population. If only one duplicate of the source is found in this sample, an 'identification' can be made.

All anthropometric measurements in the ANSUR database are reported to the nearest millimetre. IBM SPSS statistics 20 was used to search for duplicate cases within the sample. To put it simply, duplicates occur when exact values of all metric traits (traces) selected for analysis are found in more than one person. Sorting of cases was done stepwise, adding one dimension at a time to the previous dimension(s) and noting how many duplicates are found with the combination of that number of traits. This procedure was continued, increasing the number of traits until no duplicate cases were identified within the sample. For example, the first measurement analysed was bitragion submandibular arc, the number of duplicate cases identified in a sample of males was 1686. Then bizygomatic breadth was added as a second trait reducing the number of duplicate cases to 908. The third trait added was ear length, and then the number of duplicates fell down to 122. This was continued by adding ear protrusion, and then head breadth at which point no duplicates were found.

Polynomial regressions were used to study the shape of relationships between numbers of traits considered and numbers of duplicates. Those regressions allowed us to extrapolate results beyond sample sizes available.

3. Results

The sample was divided by sex to determine how many traits are needed to have no duplicate cases in each sex. The order of traits chosen was alphabetical as found in the original ANSUR files. Fig. 2 shows the percentage and number of duplicate cases for males ($N = 1774$) and females ($N = 2208$). There is a rapid decline in the number (and thus percentage) of duplicate cases with the addition of each trait to the previous trait(s). In order to have no duplicate cases in males only 5 traits are needed, whereas in females, 6 traits are needed. Thus there is little difference between the sexes in the number of traits required to reduce number of duplicate cases to zero.

Due to there being little difference in the number of traits needed to find no duplicates and thus achieve a singularity in males and females separately, the sexes were combined for all further analyses and the order of individuals was randomised.

The order of metric traits has been changed from alphabetical to largest–smallest based on the mean of each trait measured (Table 1). The order was changed to establish any differences in the number of traits needed to achieve singularity when larger metric traits were chosen first. There is a large decrease in percentage of duplicate cases (59.4%) when the second trait (bitragion submandibular arc) is added to the first trait (head circumference). In contrast, when using alphabetical order, the largest decrease between the first (bitragion submandibular arc) and second (bizygomatic breadth) traits was found in males at only 43.8% (Fig. 2). It is important to note that the overall outcome of needing 5–6 traits to achieve singularity remains the same when the order of metric traits is changed. However, the rate of decrease in duplicate cases when using metric traits in largest-to-smallest

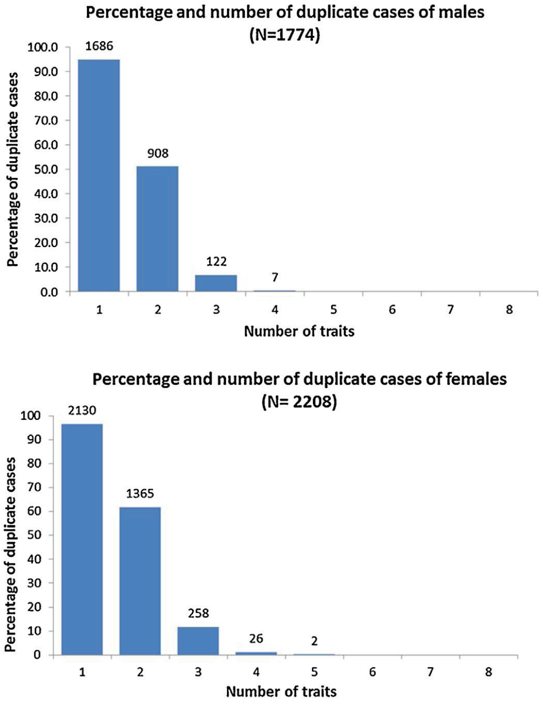


Fig. 2. The percentage and number of duplicate cases for males and females. The order of metric traits chosen is alphabetical.

Table 1
shows the metric traits used with their means for males and females combined, the order of the list changes throughout analysis (as discussed) to establish any effect the order of traits may have on the outcome.

Metric traits	Mean
Bitragion submandibular arc	288.6
Bizygomatic breadth	135.4
Ear length	62.0
Ear protrusion	22.9
Head breadth	147.7
Head circumference	555.8
Head length	191.6
Interpupillary distance	63.4

Note: this list is in alphabetical order as presented in the ANSUR database. The order of metric traits has been changed throughout the analyses as indicated.

order is faster, especially with the combination of two traits (Fig. 3).

The numbers of traits needed to achieve singularity was determined for sample sizes beginning at 200 subjects then adding 300, followed by adding 500 repeatedly at 1000, 1500, etc. until all subjects were included. This was conducted on the list of traits ranging from largest to smallest and then from smallest to largest as shown in Fig. 4. The average and maximum number of traits needed to achieve singularity when the list is ordered largest to smallest is 5. The average number of traits needed to achieve singularity when the list is ordered smallest to largest is 6 with a maximum of 7. A smaller number of traits is needed to achieve singularity when the traits are larger in dimension since they have a greater range of numerical values. However, the average number of traits (5–6) needed to achieve singularity still remains consistent, no matter in what order the traits are included into calculation. The polynomial regression equation (Fig. 4) indicates that approximately one more trait is needed for every 1000 people added to the sample in order to reduce the number of duplicate cases to zero. Polynomial regression was used to predict the number of traits needed in a sample of specific size exceeding that of the ANSUR database (Table 2). The number of possible combinations for each set of metric traits was calculated using a stepwise approach. The traits labelled 1–8 are seen here combined into sets in the order largest to smallest based on their means (Table 1). The number of units (range) is the number of possible metric values (mm) for a given metric trait, i.e. the difference between the minimum and the maximum size of the trait in full millimetres.

Probabilities of sets of traits to occur together are difficult to calculate because of the intercorrelation between them. If there is a correlation between two traits then they cannot be viewed independently of one another and their joint probability of occurrence cannot be calculated by simple multiplication of individual probabilities. This is the case with dimensions of the human face. Adjusted *R* squared values (or correlation coefficients) were calculated for each set in a stepwise approach by SPSS version 20. The *R* squared value was subtracted from one ($1-R^2$) to determine parts of traits that are not correlated and thus can be combined randomly. The total number of units possible in each set is a product of numbers of units in traits of the set reduced by their intercorrelation. The number of combinations is calculated by multiplying the $1-R^2$ value by the number of units in the first trait, then number of units in the second trait, etc. The addition of each new trait to a set increases the number of combinations as long as the trait is not completely correlated with others, and thus decreases the possibility of finding another individual with the same set of traits.

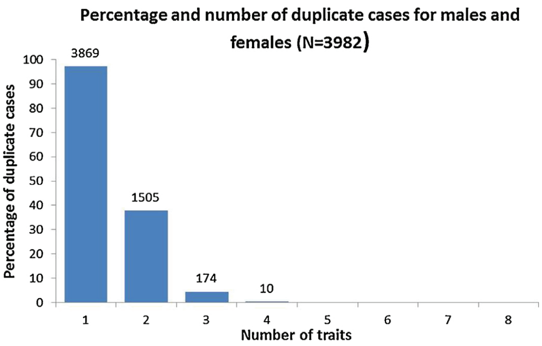


Fig. 3. The percentage and number of duplicate cases for males and females. The order of metric traits chosen from largest to smallest.

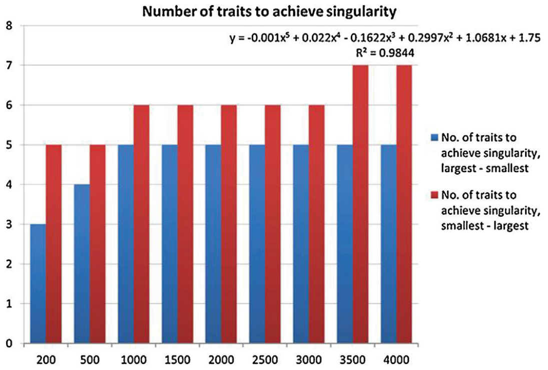


Fig. 4. The number of traits needed to have no duplicates, the list of traits is ordered largest to smallest and smallest to largest.

4. Discussion

With the increased use of video surveillance systems for identification, using the face as an example to illustrate the method proposed in this paper seemed only fitting. Many people identify individuals by the face as it is believed to be the most variable part of the human body. The findings presented in this paper have a direct application to facial analysis by providing the probabilities of finding two individuals with the exact same facial metric characteristics. As a general rule the combination of traits is the key to reducing the number of duplicates within a specified population. Depending on the exact variation of each trait, and its intercorrelation with other traits, the number of traits needed to find no duplicates and thus achieve singularity varies. This number also increases with the target population size. The larger the range of each trait, the more possible combinations per set of traits and thus the faster a result of singularity is achieved. More variable, i.e. larger traits should always be chosen first, especially in sample sizes lower than 500 individuals. In cases of applying this theory to identification from images, the traits chosen depend on the quality of the images, the angle of the cameras, whether or not the person is covering their face and many more factors. Each case is different and must be treated accordingly. However, the general guidelines still apply and the theory remains the same. If, for whatever reasons, singularity is not achieved with 8 traits, simply adding more traits will increase the probability of achieving singularity by decreasing the probabilities of finding duplicates.

Table 2
The calculation of probability of finding another duplicate within the sample, using eight metric traits, their respective ranges and R^2 values.

Traits	Number of units (range) in mm	R^2 correlation with preceding units	1-R^2	Multiplication of units	Number of combinations (1/p)	p
1	127			127	127	0.01
1, 2	138	0.415	0.585	17,526	10,253	0.0001
1, 2, 3	62	0.780	0.220	1,086,612	239,055	0.000004
1, 2, 3, 4	47	0.638	0.362	51,070,764	18,487,617	0.00000005
1, 2, 3, 4, 5	44	0.736	0.264	2,247,113,616	593,237,995	0.000000002
1, 2, 3, 4, 5, 6	26	0.380	0.620	58,424,954,016	36,223,471,490	0.00000000003
1, 2, 3, 4, 5, 6, 7	35	0.336	0.664	2,044,873,390,560	1,357,795,931,332	0.0000000000007
1, 2, 3, 4, 5, 6, 7, 8	27	0.229	0.771	55,211,581,545,120	42,568,129,371,288	0.00000000000002

When comparing metric traits with traditional descriptive traits (which use categorical scales) metric traits have a significantly larger range of values since they include more intervals (e.g. millimetres) than categorical scales. This fine gradation decreases the chances of finding a duplicate. Goldstein et al. [46] used descriptive traits with a maximum range of 5 categories per trait. Goldstein et al.'s study was able to isolate an individual from a set of 256 photographs. Not to complicate their analysis too much the authors assumed no correlation between traits, however they realised that at least some of them are correlated. Despite the simplicity of their method, Goldstein et al. [46] showed that in order to isolate an individual from a predicted population size of 10^7 it is sufficient to use 16 traits each with 5 categories. Although forensic sciences shift towards the use of traditional metric analysis for comparisons of images, the use of descriptives, however debated, is not without merit. Descriptive traits can be quantified to increase reliability. In Goldstein et al.'s case, only a limited number of categories per descriptive trait was used, thus the limited range produced inferior results. However, descriptive traits can have a significantly large range. For example, a mole above an individual's lip can be measured. Its height, width, diameter and location compared to anatomical landmarks can be calculated. Its colour can be measured metrically as a wavelength of light reflected from its surface. It may have an uneven shape that yields it to be measured in parts. Each case is different and each descriptive trait can be measured in endless ways that apply to a specific case. The same rules that apply to traditional metrics apply to descriptives, the larger the range the less chance of finding two individuals who match. However, in the case of descriptive traits, knowing the population frequencies of the occurrence of the traits would be useful in calculating probabilities.

In a study conducted by Kleinberg et al. [33], it was found that proportions of 4 anthropometric measurements taken from photographs were not sufficient to make significant positive identification of 80 individuals. From this study it was deemed that 'anthropometry failed as an identification technique'. It should be discussed that in this study there were a limited number of measurements being taken and each of these measurements was on a very small area of the face. Two of the measurements included in the four were essentially the same measurement, the distance between the ectocanthions and the stomion on both the right and left side of the face. The variability of the measurements is very limited, therefore the ability to match an individual would have been limited. This study is discussed to reinforce the importance of variability and size of the traits chosen, the larger the size and variability the higher the probability of achieving singularity.

Here we have considered correlation of traits and used only the part of each trait which varies independently from other traits. By using traits measured in interval scales and considering inter-correlations among traits we are ensuring that the method is presenting the biological variation between individuals as accurately as possible. The accuracy with which traits are measured will also affect the results. When measurement errors are known, ranges of variation should be adjusted to take errors into account. For example, if the measurement error is 3 mm, then the distance of 33 mm will legitimately be a duplicate of any distance between 30 mm and 36 mm. Size of errors depends on the quality of images measured, ability to locate points between which measurements are taken and the accuracy of measuring instruments. These factors may differ from case to case and need to be discussed when presenting results of a particular case. A study conducted by Cummaudo et al. [34] it was found that anthropometric landmarks and thus the distances between them are less reliable when taken from 2D photographs. This was also the case in a study conducted by Farkas [35]. Interobserver errors of measurements taken directly on the participants by well-trained measurers that we used in this paper were small and can be ignored for purposes of demonstration of the general principle of the proposed method. Were the method used in a particular case, errors of point location and measurements would have to be assed and taken into account when searching for possible duplicates.

In a court of law, the main question to be answered is, 'what are the chances that two sources share the same trace'. In DNA analysis, the evidential weight of a match between crime stain profile (trace) and a suspect (source) is quantified by the 'match probability' [36]. In many cases the match probability is much smaller than the probability of presence of one individual in a population which the trace originated from, with the match probability going into the inverse of millions. Table 2 shows the metric trait version of a match probability represented by 'p', the probability of finding a match of the trace to more than one individual (source). In our sample, a combination of four metric traits already gives a probability lower than that found by Goldstein et al. [46], 10^{-8}, while a combination of eight traits lowers the probability to 10^{-14}, this is comparable with DNA [36]. A fragment of the DNA molecule contains four different nucleotides that can occur in different combinations. The longer the fragment, the lower the probability that another fragment with the same pattern of nucleotides can be found. Like DNA the more information extracted from the trace, the lower the probability of finding a duplicate.

The current paper uses facial analysis as an example to support the theory of singularity, however, this theory in no way is limited to facial analysis. The concept of singularity and the method presented can be applied to all fields of forensic sciences which aim to match a trace to a source. Many biological examples have been presented in this paper, however. the concept can also be applied to matching a trace with a source left by an object, for example, the study of ballistics [37], gunshot residues [38] tool marks [39] tire prints [40] cut marks and tool marks [41] glass fragments [42] and any other pattern matching of everyday objects [43].

5. Conclusions

This paper introduces the concept of singularity to forensic identification. Within the specified population of the ANSUR database, no two individual faces matched one another on combinations of 5–8 metric traits. Probabilities of finding a duplicate of a face characterised by 8 traits exceeded inverse of the total population of the Earth making metric identification of faces as reliable as that of DNA. The same concept can be applied to identification based on measurements of human bodies or any traces of any objects.

References

[1] M. Page, J. Taylor, M. Blenkin, Uniqueness in the forensic identification sciences – fact or fiction? Forensic Sci. Int. 206 (2011) 12–18.
[2] S. Cole, Forensics without uniqueness, conclusions without individualization: the new epistemology of forensic identification, Law Probab. Risk 8 (2009) 233–255.
[3] A.J. Jeffreys, V. Wilson, S.L. Thein, 'Individual specific' fingerprints of humans DNA, Nature 316 (1985) 76–79.
[4] A.K. Jain, L. Hong, S. Pankanti, R. Bolle, An identity authentication system using fingerprints, Proc. IEEE 85 (1997) 1365–1388.
[5] A.K. Jain, S. Prabhakar, S. Pankanti, On the similarity of identical twin fingerprints, Pattern Recognit. 35 (2002) 2653–2663.
[6] A. Mahaja, A. Bartra, B. Khurana, J. Kaur, Role of bitemark analysis in identification of a person, Glob. J. Med. Public Health 1 (1) (2012) 56–59.
[7] R.T. Oatess, Elbow print identification, J. Forensic Identif. 50 (2000) 132–137.
[8] L. Meijerman, S. Sholl, F. De Conti, M. Giacon, C. Van Der Lugt, A. Drusini, P. Vanezis, G. Maat, Exploratory study on classification and individualisation of earprints, Forensic Sci. Int. 140 (1) (2004) 91–99.
[9] G. Mishra, K. Ranganathan, T. Saraswathi, Study of lip prints, J. Forensic Dent. Sci. 1 (2009) 28.
[10] I. Bouchrika, M. Goffredo, J. Carter, M. Nixon, On using gait in forensic biometrics, J. Forensic Sci. 56 (4) (2011) 882–889.
[11] A. Crane, Does the amount of handwriting on a cheque constitute a 'reasonable amount of sample'? Can. Soc. Forensic Sci. J. 32 (1) (1999) 39–45.
[12] M. Henneberg, Facial Mapping, Body Mapping and the Duties of An Expert Witness, 2007 http://www.lawlink.nsw.gov.au/lawlink/pdo/ll_pdo.nsf/pages/PDO_facialmapping.
[13] M. Henneberg, Expert witness in a courtroom: the Australian experience, in: M. Oxenham (Ed.), Forensic Approaches to Death Disaster and Abuse, Australian Academic Press, Queensland, 2008, pp. 307–319.
[14] F. Rosing, Identification von Personen auf Bildern, Verlag C.H. Beck, München, 2006.
[15] G. Edmond, Specialised knowledge, the exclusionary discretions and reliability: reassessing incriminating expert opinion evidence, Univ. New South Wales Law J. 31 (1) (2008) 1–55.
[16] G. Edmond, K. Biber, R. Kemp, G. Porter, Law's looking glass: expert identification evidence derived from photographic and video images, Curr. Issues Crim. Justice 20 (3) (2009) 337–376.
[17] G. Edmond, Impartiality, efficiency or reliability? A critical response to expert evidence law and procedure in Australia, Aust. J. Forensic Sci. 42 (2) (2010) 83–99.
[18] T. Scoleri, M. Henneberg, View-independent prediction of body dimensions in crowded environments, in: International Conference on Digital Image Computing Techniques and Applications (2012: Fremantle) DICTA, 2012.
[19] T. Scoleri, T. Lucas, M. Henneberg, Effect of garments on photoanthropometry of body parts: application of stature estimation, Forensic Sci. Int. 237 (2014) 1–12.
[20] K. Shi, A. Samal, D. Marx, How effective are landmarks and their geometry for face recognition? Comput. Vis. Image Underst. 102 (2006) 117–133.
[21] A. Burton, S. Wilson, M. Cowan, V. Bruce, Face recognition in poor-quality video: evidence from security surveillance, Psychol. Sci. 10 (1999) 243–248.
[22] R. Jafri, H. Arabnia, A survey of face recognition techniques, J. Inf. Process. Syst. 5 (2) (2009) 41–68.
[23] Chen, S. Mau, S. Harandi, M. Sanderson, C. Bigdeli, A. B Lovell, Face recognition from still images to video sequences; a local-feature-based framework, J. Image Video Process. 2011 (2010) 1–14.
[24] C. Cattaneo, Z. Obertová, M. Ratnayake, L. Marasciuolo, J. Tutkuviene, P. Poppa, D. Gibelli, P. Gabriel, S. Ritz-Timme, Can facial proportions taken from images be of use for ageing in case of suspected child pornography? A pilot study, Int. J. Legal Med. 126 (2012) 139–144.
[25] B. Klare, A.K. Jain, On a taxonomy of facial features, in: 2010 Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS), IEEE, 2010, pp. 1–8.
[26] S. Ritz-Timme, P. Gabriel, J. Tutkuviene, P. Poppa, Z. Obertová, D. Gibelli, D. De Angelis, M. Ratnayake, R. Rizgeliene, A. Barkus, C. Cattaneo, Metric and morphological assessment of facial features: a study of three European populations, Forensic Sci. Int. 207 (2011) 239e1–239e8.
[27] M. Turk, Pentland A., Eigenfaces for recognition, Cogn. Neurosci. 3 (1991) 72–86.
[28] N.A. Spaun, Forensic biometrics from images and video at the Federal Bureau of investigation, in: Biometrics: Theory Applications and Systems (BTAS), 2007, 1–3.
[29] N. Spaun, Facial comparisons by subject matter experts: their role in biometrics and their training, in: International Conference on Advances in Biometrics, 2009, 161–168.
[30] C.E. Clauser, I.O. Tebbetts, B. Bradtmiller, J.T. McConville, C.C. Gordon, ANSUR Database: Measurer's Handbook: US Army Anthropometric Survey 1987–1988. Technical Report NATICK/TR-88/043, 1988.
[31] M. Henneberg, The Illusive concept of human variation: Thirty years of teaching biological anthropology on four continents, in: G. Štrkalj (Ed.), Teaching Human Variation, Nova Science, New York, 2010, pp. 33–43.
[32] L.L. Cavalli-Sforza, W.F. Bodmer, The Genetics of Human Population, W.H. Freeman and Company, San Francisco, 1971.
[33] K. Kleinberg, P. Vanezis, A.M. Burton, Failure of anthropometry as a facial identification technique using high quality photographs, J. Forensic Sci. 52 (4) (2007) 779–783.
[34] M. Cummaudo, M. Guerzoni, L. Marasciuolo, D. Gibelli, A. Cigada, Z. Obertová, M. Ratnayake, P. Poppa, P. Gabriel, S. Ritz-Timme, C. Cattaneo, Pitfalls at the root of facial assessment on photographs: a quantitative study of accuracy in positioning of facial landmarks, Int. J. Legal Med. 127 (2013) 699–706.
[35] L.G. Farkas, Anthropometry of the Head and Face, Raven Press, New York, 1994.
[36] M.A. Jobling, P. Gill, Encoded evidence: DNA in forensic analysis, Nat. Rev. Genet. 5 (10) (2004) 739–751.
[37] E. Springer, Toolmark examinations – a review of its development in the literature, J. Forensic Sci. 40 (6) (1995) 964–968.
[38] D. Gibelli, A. Brandone, S. Andreola, D. Porta, E. Giudici, M. Grande, C. Cattaneo, Macroscopic, microscopic and chemical assessment of gunshot lesions on decomposed pig skin, J. Forensic Sci. 55 (4) (2010) 1092–1097.
[39] F.H. Cassidy, Examination of tool marks from sequentially manufactured tongue-and-groove pliers, J. Forensic Sci. 25 (4) (1980) 796–809.
[40] P. McDonald, Tire Imprinting Evidence, CRC Press, Florida, 1993.
[41] K.J. Reichs, Postmortem dismemberment: recovery, analysis and interpretation, in: Forensic Osteology: Advances in the Identification of Human Remains, 1998, 353–388.
[42] E.M. Rodriguez-Celis, I.B. Gornushkin, U.M. Heitmann, J.R. Almirall, B.W. Smith, J.D. Winefordner, N. Omenetto, Laser induced breakdown spectroscopy as a tool for discrimination of glass for forensic application, Anal. Bioanal. Chem. 391 (2008) 1961–1968.
[43] P.T. Jayaprakash, Practical relevance of pattern uniqueness in forensic science, Forensic Sci. Int. 231 (2013) 1–16.
[44] C.C. Gordon, B. Bradtmiller, T. Churchill, C.E. Clauser, J.T. McConville, I.O. Tebbetts, R.A. Walker, Anthropometric Survey of US Army Personnel: Methods and Summary Statistics. Technical Report NATICK/TR-89/044, United States Army Natick Research, Development and Engineering Center, Natick, Massachusetts, 1988.
[45] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.
[46] A.J. Goldstein, L.D. Harmon, A.B. Lesk, Identification of human faces, Proc. IEEE 59 (5) (1971) 748–760.

Appendix 3: Reprint of ‘Comparing the face to the body, which is better for identification?’

Lucas, T. & Henneberg, M. (2015). Comparing the face to the body, which is better for identification?
International Journal of Legal Medicine, 130(2), 533-540.