Forensic Science International 233 (2013) 133–139

Contents lists available at ScienceDirect

Forensic Science International

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

Preliminary performance assessment of computer automated facial approximations using computed tomography scans of living individuals

Connie L. Parks, Adam H. Richard, Keith L. Monson *

Counterterrorism and Forensic Science Research Unit, FBI Laboratory Division, 2501 Investigation Parkway, Quantico, VA 22135, United States

ARTICLE INFO

Article history:
Received 8 March 2013
Received in revised form 10 May 2013
Accepted 30 August 2013
Available online 8 September 2013

Keywords:
Forensic anthropology
Facial reconstruction
Computerized facial approximation
Resemblance

ABSTRACT

ReFace (*Reality Enhancement Facial Approximation by Computational Estimation*) is a computer-automated facial approximation application jointly developed by the Federal Bureau of Investigation and GE Global Research. The application derives a statistically based approximation of a face from a unidentified skull using a dataset of ~400 human head computer tomography (CT) scans of living adult American individuals from four ancestry groups: African, Asian, European and Hispanic (self-identified). To date only one unpublished subjective recognition study has been conducted using ReFace approximations. It indicated that approximations produced by ReFace were recognized above chance rates (10%). This preliminary study assesses: (i) the recognizability of five ReFace approximations; (ii) the recognizability of CT-derived skin surface replicas of the same individuals whose skulls were used to create the ReFace approximations; and (iii) the relationship between recognition performance and resemblance ratings of target individuals. All five skin surface replicas were recognized at rates statistically significant above chance (22–50%). Four of five ReFace approximations were recognized above chance (5–18%), although with statistical significance only at the higher rate. Such results suggest reconsideration of the usefulness of the type of output format utilized in this study, particularly in regard to facial approximations employed as a means of identifying unknown individuals.

Published by Elsevier Ireland Ltd.

1. Introduction

When presented with the skeletal remains of an unknown individual, those charged with establishing the identity of the remains often must exhaust a number of avenues before a positive identification is made. Although not a form of positive identification, facial approximation is an avenue often utilized when more traditional methods fail to produce leads. "Facial approximation" refers to the technique of developing an antemortem representation from the skeletonized skull of an individual [1–4]. Facial approximation techniques are a unique blend of both science and art that date back to the Neolithic period (9000 BC–3000 BCE). One of the first attempts to recreate the visage of a deceased individual was found in Jericho in the Jordan Valley where inhabitants of the city applied clay to the dry skulls of their dead as a symbolic form of ancestor worship [1,2,5–7]. Although crude, the Jericho death masks provide some of the earliest evidence of attempts to reconstruct faces [7]. The first modern attempt to reconstruct the visage of an individual is credited to German anatomist Wilhelm His in 1895 for his reconstruction of composer Johann Sebastian Bach [1,3–5,7,8]. His's reconstruction sparked a new interest in recreating the likenesses of individuals for a variety of purposes, including archeological investigations, historic authentications, and more recently, the identification of unknown individuals [3,5–8].

Currently, facial approximations are developed using one of three main techniques: (i) manual 2D drawing, (ii) manual 3D sculpting and (iii) 2D and 3D computer-automated modeling [1,9]. The accuracy (i.e., correct recognition of the individual to whom a skull belongs) of approximations produced by these techniques, particularly the computer-automated technique, is controversial and often challenged [4,10–13]. A number of empirical validation protocols have been developed to address this challenge [10,12,13]. Two of the more commonly used protocols are: (i) a comparison between an approximation and a pool of photographs (referred to as recognition testing) and (ii) a direct comparison between an approximation and a photograph of the individual from whose skull the approximation is constructed (referred to as resemblance ranking).

The first attempt to validate the accuracy of a facial approximation is credited to H. Von Eggeling in 1913 when he attempted to reproduce the visage of a cadaver using tissue depths

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

0379-0738/$ – see front matter. Published by Elsevier Ireland Ltd.
http://dx.doi.org/10.1016/j.forsciint.2013.08.031

and subsequent comparison to the cadaver's death mask [3,7,10]. Since Von Eggeling's time, numerous studies have been published regarding the accuracy of non-computerized facial approximations. In contrast, accuracy studies for computer-automated techniques are few [3,14].

Due to the paucity of published validation studies for computer automated facial approximation techniques, it is essential that further validation research be conducted in an effort to collect accuracy data not only for ReFace, the computer-automated approximation application utilized in this study, but for all current and future techniques. Only with such data can performance thresholds be elucidated for facial approximation tools. The preliminary study assesses: (i) the recognizability of ReFace approximations; (ii) the recognizability of CT-derived skin surface replicas of the same individuals whose skulls were used to create the ReFace approximations; and (iii) the relationship between recognition performance and resemblance ratings of target individuals.

1.1. ReFace

ReFace (*Reality Enhancement Facial Approximation by Computational Estimation*) is a computer-automated facial approximation application jointly developed by the Federal Bureau of Investigation (FBI) and General Electric (GE) Global Research. The application derives a statistically based approximation of a face from an unidentified skull using a dataset of ~400 human head computer tomography (CT) scans [15]. For a detailed explanation of ReFace foundations, see references [15–19]. ReFace is not publicly available at this time. Although objective validation studies [20,21] have been conducted using ReFace approximations, only one unpublished subjective recognition study has been completed to date [6]. That study indicated that approximations produced by ReFace were recognized 10% above chance rates, a statistic comparable to other facial approximation techniques [2,12,22].

2. Materials and methods

2.1. Dataset and image acquisition

The ReFace dataset consists of ~400 medical CT scans of living adult American individuals with equal representation of four ancestry groups: African, Asian, European, and Hispanic (all self-identified). A subset of individuals was selected from the ReFace dataset for inclusion in this study based on equivalent scan protocols and availability of consistent photos (i.e., similar quality, pose, and lighting). From this subset, further selection was based on the same sex and ancestry and comparable age and weight. Individuals with obvious craniofacial trauma were excluded. Five individuals met all of the above criteria and were selected as the final target individual sample (see Fig. 1). The scan protocol for the five selected individuals consisted of slice thickness of 2.5 mm, slice increments of 1.70 mm, resolution of 512 × 512, and pixel size of 0.488. All anonymized data were approved for use by the FBI's Institutional Review Board. Protection of the anonymity of the target individuals precludes publication of their photographs or CT-derived skin surface replicas.

The facial approximations for the target individuals were generated using ReFace, the targets being temporarily removed from the reference database for the purpose of these studies. Although ReFace offers the capability to alter the approximations by constraining for age and weight, no such constraints were applied to the approximations produced for this study. Instead, the "average" approximations produced by the application, as if the age and weight were unknown, were used. Additionally, no texture, pigmentation, hair, or other modifications were applied to the approximations (Fig. 2).

Non-target individuals were also selected from the ReFace dataset based on the same sex and ancestry as the target individual. This represents reasonable initial information, typically provided by a forensic anthropologist, with which to begin an approximation (manual or computer automated). Further selection was narrowed based on similarity of age and weight with the target individual and the availability of clear, full frontal pose photos with no facial expressions.

The skin surface replica images were obtained by importing and segmenting the CT scans of the target individuals using Mimics v.14.11 (Materialise, Ann Arbor, MI). The skin layer was segmented from the bone using the application preset soft tissue Hounsfield threshold value of −700 to +225. No further processing was performed on the segmented skin surface replicas.

2.2. Construction of recognition and resemblance tests

For the facial approximation recognition tests, photo arrays were assembled using one ReFace approximation and six photographs in full frontal orientation. No photographs were repeated in any of the six photo arrays. The six photographs included the target and five non-target individuals of the same sex and ancestry and the same approximate age and weight. The ReFace approximations and photos were standardized in terms of pose, lighting, quality, and size (90 H × 60 W mm). The photo arrays were printed in black and white on 8.5″ by 11″ HP Advanced glossy photo paper using an HP Deskjet 6940 printer. To determine the functionality of the photo arrays, four individuals, with no stake in the study and no prior knowledge of the array photos, were asked to evaluate the arrays. The individuals were asked if any photo(s) in each array was noticeably different from the other photos in the array. All four individuals indicated that none were.

For the skin surface replica recognition tests, the photo arrays were constructed following the same procedure as the ReFace approximation recognition arrays with the exception that the ReFace approximation was replaced with a CT-derived skin surface replica in full frontal orientation. For the ReFace approximation resemblance tests, the same procedure was followed as above with the exception that only the target individual and the ReFace approximation of that target were printed. The images for the resemblance tests were larger than for the recognition tests, measuring 115 H × 80 W mm. The skin surface replica resemblance tests were constructed in the same manner as the ReFace approximation resemblance tests except that the approximation was replaced with a skin surface replica image.

2.3. Participant responsibilities

Each participant viewed all five target individuals. For each target, participants were asked to perform four tasks (totaling 20 tasks):

1. View a ReFace approximation and an array of photographs and identify the individual in the array represented by the approximation (recognition test). Participants had the option to judge that there was no match and the individual was not present.
2. View a skin surface replica and an array of photographs and identify the individual in the array represented by the skin surface replica (recognition test). Participants had the option to judge that there was no match and the individual was not present.
3. View a one-to-one comparison of a ReFace approximation and a photo of the individual from whom the approximation was created and rate how closely the approximation resembles the individual (resemblance test).
4. View a one-to-one comparison of a skin surface replica and a photo of the individual from whom the model was created and rate how closely the skin surface replica resembles the individual (resemblance test).

2.4. Study duration and test material randomization

The current study was conducted over a 3 day period, presenting the possibility of potentially detrimental post-participation discussion with potential assessors. To control for this, two sets of test materials were produced (set 1 and set 2). The order in which the photo arrays were presented and the photo positions within each array were altered between sets. In other words, a correct identification in set 1 would not be correct in set 2. The sets were given to participants randomly.

2.5. Assessment procedure

The recognition and resemblance tests were administered using unfamiliar assessors (no familiarity between assessors and target individuals) in a simultaneous array presentation. A total of 119 FBI Laboratory employees voluntarily participated in the study (39 males and 80 females; range 23–63 years). Only 16 participants were allowed into the testing area at the same time. After one participant completed the study, another participant was given testing materials and allowed to enter the testing area. As part of the study, participants were asked to read a brief overview of facial approximation construction and validation methods (see Appendix A). Participants were instructed to notify the study administrator if they recognized any individuals pictured in the photo arrays, so as to be removed from participation. No time limit was given for completing the test materials.

Each participant received a packet containing four envelopes:

Envelope 1. ReFace approximation recognition materials with answer sheet.

Envelope 2. Skin surface replica recognition materials with answer sheet.