Automated Facial Recognition of Computed Tomography-Derived Facial Images: Patient Privacy Implications

Connie L. Parks¹ · Keith L. Monson²

Published online: 26 December 2016
© Society for Imaging Informatics in Medicine 2016

Abstract The recognizability of facial images extracted from publically available medical scans raises patient privacy concerns. This study examined how accurately facial images extracted from computed tomography (CT) scans are objectively matched with corresponding photographs of the scanned individuals. The test subjects were 128 adult Americans ranging in age from 18 to 60 years, representing both sexes and three self-identified population (ancestral descent) groups (African, European, and Hispanic). Using facial recognition software, the 2D images of the extracted facial models were compared for matches against five differently sized photo galleries. Depending on the scanning protocol and gallery size, in 6–61 % of the cases, a correct life photo match for a CT-derived facial image was the top ranked image in the generated candidate lists, even when blind searching in excess of 100,000 images. In 31–91 % of the cases, a correct match was located within the top 50 images. Few significant differences ($p > 0.05$) in match rates were observed between the sexes or across the three age cohorts. Highly significant differences ($p < 0.01$) were, however, observed across the three ancestral cohorts and between the two CT scanning protocols. Results suggest that the probability of a match between a facial image extracted from a medical scan and a photograph of the individual is moderately high. The facial image data inherent in commonly employed medical imaging modalities may need

to consider a potentially identifiable form of “comparable” facial imagery and protected as such under patient privacy legislation.

Keywords Facial recognition · Patient privacy · HIPAA · Computed tomography (CT)

Introduction

Computed tomography (CT) is a common radiographic modality routinely, and increasingly, employed in diagnostic medicine [1–5]. When amassed in institutional or publically accessible repositories, radiographic data provide an invaluable resource to the research, educational, and scientific communities [6–9]. Institutions submitting radiographic data to public repositories are mandated by the Health Information and Portability Act of 1996 (HIPAA) to anonymize the data by removing all protected health (PHI) and personally identifiable information (PII) prior to submission [10]. HIPAA specifically addresses 18 PHI and PII identifiers summarily defined as information that can, either alone or in conjunction with other data, lead to the identity of a particular individual. Full-face photographs and any “comparable” images are two identifiers specifically protected under HIPAA provisions [11]. HIPAA legislation does not, however, precisely define what constitutes a comparable facial image, particularly in regards to the facial data available in specific medical imaging modalities such as computed tomography.

In addition to capturing digital images of internal body features (e.g., skeletal features), certain medical imaging modalities (e.g., CT) also capture the external skin surface. If the medical scan is of the head and neck region, the possessor of

✉ Keith L. Monson
Keith.Monson@ic.fbi.gov

¹ Counterterrorism and Forensic Science Research Unit, Visiting Scientist Program, FBI Laboratory Division, 2501 Investigation Parkway, Quantico, VA 22135, USA

² Counterterrorism and Forensic Science Research Unit, FBI Laboratory Division, 2501 Investigation Parkway, Quantico, VA 22135, USA

the scan essentially holds all the requisite data to extract a “sculpted rendition” of an individual’s face [12–15]. Previously not widely available or affordable to the average computer user, the software tools necessary to extract such images from medical scans are now numerous and easily accessible, often at no cost to the end user, e.g., [16–18]. The ability to readily acquire medical scans containing extractable facial data raises concerns regarding the potential for personal identification from, or malicious exploitation of, facial skin surface images extracted from medical scans. Researchers and privacy advocates have expressed concerns and question whether or not these types of “reconstructed” facial images should also be considered PHI/PII and explicitly protected under HIPAA [15, 19–25].

Few studies exist examining the recognizability of facial images reconstructed from medical scans, by either human reviewers or objectively via the use of automated facial recognition systems [15, 19, 20, 26]. The only known research examining the automated recognition of reconstructed facial images was conducted by Mazura et al. [20], who, using a freely available facial recognition system [27] and a pool of $n = 179$ photos, report a 27.5 % match between the CT-derived facial images and corresponding photos of the same individuals ($n = 29$). Facial recognition systems employ an array of complex algorithms to compare a facial image of interest (*probe*) against a reference set of images (*gallery*). Depending on the operational implementation, a single potential match (1:1 verification) or multiple potential matches (1:N candidate lists) are produced. Multiple 1:N matches are selected from the gallery images based on user-defined search parameters and ranked according to application-specific scoring criteria, effectively providing a prioritized list of possible matches, with the top ranked image representing a potential positive identification.

Given the public availability of CT repositories, image extraction software, and facial recognition systems, it is posited that individual identification from CT-extracted facial images is feasible. The purpose of this research is objective examination of this hypothesis with the intent to provide preliminary data to inform concerns regarding the recognizability of facial images derived from medical scans. Utilizing a commercial facial recognition system [28] this study objectively examined how accurately facial images reconstructed from computed tomography scans were matched with corresponding photographs of the scanned individuals. The results of this study will (i) contribute to existing literature examining the recognizability of facial images derived from medical scans, (ii) expand patient awareness for informed consent, (iii) inform policy regarding application of medical data anonymization and or obscuration techniques, and (iv) stimulate discourse intended to modernize and broaden the scope of patient privacy legislation.

Methods

The research model employed in this study has been reported by Parks and Monson [29] and is only briefly discussed here for the purpose of pointing out certain methodological differences. Given that the ability of the facial recognition system employed in this research to successfully process unconventional, non-photographic probes has been established in prior research, [29, 30] no internal control was conducted. In contrast to the methodology of Parks and Monson: [29] (i) only blind searches against the entirety of the gallery(ies) were conducted, (ii) no image quality enhancements were applied (e.g., contrast), and (iii) the probe set consisted of a single CT-derived facial image for each test subject.

Galleries

The original image gallery ($g = 6159$) previously constructed by Parks and Monson [29] was used in this study. In addition to the original gallery, the authors of the current paper also constructed four additional galleries, one each $g = 12,000$; 25,000; 50,000; and 108,000. All galleries were assembled from the same parent source as [29]. The galleries were constructed in such a manner that each gallery was approximately double in size relative to the preceding smaller gallery. Further, each subsequently constructed larger gallery contained the prior smaller gallery images. For example, the $g = 12,000$ gallery contained the $g = 6159$ images, as well as approximately 6000 randomly selected additional images. Detailed gallery image properties (e.g., resolution and composition, inter-pupillary distance, pose, illumination, and expression) for the $g = 6159$ are detailed in [29]. Although the original $g = 6159$ was highly standardized with all images individually vetted for image quality, demographics, and adherence to biometric standards (e.g., ANSI/NIST-ITL 1-2000), the images in the additional four galleries were not vetted for these attributes. Given the random selection and absence of image vetting, the quality and demographic distribution of the four additional galleries was not confirmed.

Test Subjects and Probe Set

The test subjects ($t = 128$) in this study were a subset of the original sample ($t = 159$) compiled by Parks and Monson [29]. The test subjects were adult Americans ranging in age from 18 to 60 years with representation of both sexes and three self-identified population (ancestral descent) groups (African, European, and Hispanic). The test subjects’ CT scans, from which the facial images were extracted, comprised two scanning protocols. The parameters for each protocol are detailed in Table 1 and are hereafter referred to by the protocol column labels, SP3.5 and SP6.0. Following segmentation of the facial data from the CT scan (see [26] for segmentation details), a

Table 1 Scanning protocol parameters

Variable	SP3.5	SP6.0
Test subjects	$N = 33$	$N = 95$
Scanner	GE Imatron C-150HR	GE Imatron e-speed
kV	130	140
mA	247.60	299.70
Algorithm	Normal	Normal
Reduction	1	1
Orientation	RAB	RAB
Field of view (cm)	30.00	26.00
Gantry tilt	1	0
Slice thickness (mm)	3.50	6.00
Slice increment (mm)	1.70	3.00
Pixel size (mm)	0.586	0.508
Slice width (pixel)	512	512
Slice height (pixel)	518	512

single, Frankfurt-aligned 2D image was exported for each test subject and subsequently compiled as the probe set ($p = 128$) for this study. The only modification applied to the CT-derived facial images was an automated cropping procedure intended to remove excess background and produce a “head and shoulders” image composition (see [26] for description of the auto crop procedure). All probe images were captured in a grayscale JPG file format.

The Institutional Review Boards of collaborating medical institutions and the FBI approved the collection and use of the images employed in this study. Protection of the anonymity of the living test subjects precludes publication of either life photos or CT-derived facial images. The images depicted in Fig. 1 are artist-rendered composites compiled from the CT scans of multiple individuals and do not reflect actual living persons. These images are provided to demonstrate the appearance of the CT-derived images used in this study.

Rank Classes

The primary goal of a 1: N identification task within a facial recognition scenario is not that a test subject’s corresponding

gallery photo is always the top image in a candidate list, but rather that a test subject’s gallery photo is one of the top k images [31–34]. Given this definition, a correct match in this study indicates that a candidate list included a test subject’s corresponding life photo within the top k images. The set of top k images is referred to as a *rank class* (R_k) and includes all images that ranked from 1 to k , inclusively. The four rank classes defined for this study are (i) R_1 , (ii) R_{10} , (iii) R_{25} , and (iv) R_{50} .

Performance Evaluation and Statistical Analysis

Two performance measures are provided: (i) *identification* (R_1) and *candidate list inclusion* rates (R_{10} , R_{25} , R_{50}) for each gallery, demographic cohort, and scan protocol, and (ii) statistical analysis of cohort relationships across the five galleries, three ancestral and age groups, and between the two scan protocols. The number of correct matches at the R_1 rank class is referred to in this study as an identification rate, as these matches represent potential positive identifications [33, 35]. Contrastingly, the number of correct matches within rank classes other than R_1 is referred to as a candidate list inclusion rate. Both the identification and candidate list inclusion rates are expressed as the fraction of total test subjects (t) who were identified within a specific rank class (R_k) and estimates the probability of obtaining a correct match within the top k images in a candidate list. For example, if in a blind search $R_{25} = 77$ and $t = 128$, the candidate list inclusion probability is R_{25}/t , or 0.60, for this specific search type and rank class.

This study was conducted in a repeated measures design with five treatments (galleries), four treatment levels (rank classes), and one dichotomous variable (correct or incorrect R_k placement). A Cochran’s Q test indicated that highly significantly different ($p < 0.01$) match rates existed across the galleries and rank classes. For further clarification, a series of McNemar’s tests was conducted to examine performance differences between gallery and rank class pairs. Fisher’s Exact tests were used to analyze performance between non-matched demographic cohorts (e.g., male vs. female). A sequential Holm-Sidak procedure [36] was employed to control for the possible effect of simultaneous comparisons (>2 groups).

Fig. 1 CT-derived facial image examples. Note: These images are artist-rendered composites from multiple CT scans and do not reflect actual living persons

Statistical significance was assessed at $\alpha = 0.05$. Data were analyzed using SPSS 16.0 (Chicago, IL) and Microsoft Excel 2007 (Redmond, WA).

Results and Discussion

Overall and demographic cohort performance metrics are extensively detailed in Table 2. Consistently observed in this study, and similar to previously published literature [29, 30, 33, 35], was an inverse relationship between identification and candidate list inclusion rates and (i) gallery size and (ii) rank class. Given this empirically observed relationship, the large number of performance metrics calculated, and to facilitate manuscript brevity, results are presented only for the R_1 rank class. The R_1 rank class represents potential positive identifications [33, 35] and is, therefore, the most relevant to the purpose of this study. However, assuming that some operational cases (e.g., unidentified persons investigations) may involve an equally weighted assessment of the top k images of a candidate list, other noteworthy R_k results are also presented (see Table 2 for metrics not expressly discussed). Performance measures (in terms of gallery size and rank class) are presented for performance: (a) overall (i.e., results from both scan protocols combined), (b) of the facial images derived from SP3.5 only, (c) of the facial images derived from SP6.0 only, and (d) *between* the two scan protocols. The results of the McNemar's pairwise tests are also briefly addressed. In addition, sub-discrimination of some of the demographic cohorts (e.g., males further subdivided by ancestry and or age) presented in Table 2 resulted in sample sizes too small for meaningful analysis (e.g., $n < 5$). Although performance metrics for these cohorts are presented for completeness in Table 2, discussion regarding these cohorts is not included. Cautious interpretation of the results for these under-represented cohorts is advised.

Overall Performance

Overall R_1 identification rates ranged from 14 to 27 % for the combined scan protocols, depending on the gallery employed (Table 2). Comparable, and not significantly different (all $p > 0.16$), R_1 identification rates were also observed for the male (15–28 %), female (13–27 %), and 18–29-year-old (17–37 %) cohorts. Although also not significantly different (all $p > 0.06$) than the overall rates, slightly lower R_1 identification rates were observed for the African (8–19 %), Hispanic (8–18 %), 30–49-year-old (10–13 %), and ≥ 50 -year-old (10 %) cohorts. In contrast, the Caucasian cohort exhibited significantly higher (all $p < 0.01$) R_1 identification rates (40–64 %). Noteworthy is that, depending on gallery size, in 43–67 % of the cases, a correct life photo match for a CT-derived facial image was located within the top 50 images (R_{50}) of the

generated candidate lists, even when a gallery in excess of 100,000 images was employed. In this study, the top 50 images represented < 1 % of the total images contained within each of the galleries. These results suggest that the probability of a match between the facial image extracted from a medical scan and a photograph of the individual is moderately high within as little as the top 1 % of gallery images presented as potential matches, irrespective of gallery size.

Examination of cohort relationships indicated that few significant performance differences (majority of $p > 0.05$) exist between the sexes or across the three age cohorts, irrespective of R_k or gallery size. Highly significant differences were, however, consistently observed between the European cohort and the African and Hispanic cohorts across most R_k and gallery sizes (majority of $p < 0.05$). Although additional significant differences were also observed for demographic cohorts sub-discriminated by ancestry, sex, and or age, most were inconsistent across R_k and gallery size, generally exhibiting non-significance with the use of larger galleries (e.g., $g = 50,000$ and $g = 108,000$).

Scan Protocol SP3.5 Performance

R_1 identification rates for scan protocol SP3.5 ranged from 36 to 61 %, depending on the gallery employed (Table 2). Comparable, and not significantly different (all $p > 0.49$), R_1 identification rates were also observed for the male (33–52 %) and 18–29-year-old (31–59 %) cohorts. Although also not significantly different (all $p > 0.58$) than the overall rates for this protocol, slightly higher R_1 identification rates were observed for the female (42–75 %) and Caucasian (40–64 %) cohorts. The African cohort exhibited the poorest performance (25–50 %) observed in SP3.5, although again, not significantly different (all $p > 0.69$) than the overall rates. Of note is that, depending on the gallery size employed, in 79–91 % of the cases, a correct life photo match for a CT-derived facial image was located within the top 50 images (R_{50}) of the generated candidate lists for this protocol. Examination of cohort relationships for SP3.5 indicated that no significant performance differences (all $p > 0.15$) exist between the sexes or across the ancestral and age cohorts eligible (i.e., $n \geq 5$) for examination, irrespective of R_k , gallery size, or cohort sub-discrimination.

Scan Protocol SP6.0 Performance

R_1 identification rates for scan protocol SP6.0 ranged from 6 to 16 %, depending on the gallery employed (Table 2). Comparable, and not significantly different (all $p > 0.72$), R_1 identification rates were also observed for the male (6–17 %), female (6–15 %), African (4–13 %), and Hispanic (8–18 %) cohorts. Although also not significantly different (all $p > 0.26$) than the overall rates for this protocol, slightly higher R_1 identification rates were observed for the 18–29-year-old cohort

Table 2 Overall and demographic cohort identification (R_1) and candidate list inclusion (R_{10} , R_{25} , R_{50}) rates by CT scan protocol (SP3.5 and SP6.0), rank class (R_k), and gallery (g) size

		SP3.5 and SP6.0										SP3.5				
		R_k	t	$(g = 6159)$	$(g = 12,000)$	$(g = 25,000)$	$(g = 50,000)$	$(g = 108,000)$	t	$(g = 6159)$	$(g = 12,000)$	$(g = 25,000)$	$(g = 50,000)$	$(g = 108,000)$		
Overall		R_1	128	0.27	0.27	0.22	0.22	0.14	33	0.61	0.61	0.52	0.52	0.36		
		R_{10}		0.51	0.45	0.38	0.33	0.30		0.82	0.76	0.70	0.61	0.55		
		R_{25}		0.59	0.54	0.50	0.43	0.34		0.88	0.85	0.82	0.79	0.64		
		R_{50}		0.67	0.59	0.55	0.51	0.43		0.91	0.88	0.88	0.82	0.79		
Cohorts																
Male		R_1	68	0.28	0.26	0.21	0.21	0.15	21	0.52	0.52	0.43	0.43	0.33		
		R_{10}		0.56	0.53	0.43	0.35	0.31		0.81	0.76	0.67	0.52	0.48		
		R_{25}		0.60	0.56	0.54	0.51	0.38		0.86	0.81	0.81	0.81	0.62		
		R_{50}		0.69	0.60	0.59	0.54	0.51		0.86	0.86	0.86	0.81	0.81		
Female		R_1	60	0.27	0.27	0.23	0.23	0.13	12	0.75	0.75	0.67	0.67	0.42		
		R_{10}		0.45	0.35	0.33	0.30	0.28		0.83	0.75	0.75	0.75	0.67		
		R_{25}		0.58	0.52	0.45	0.33	0.28		0.92	0.92	0.83	0.75	0.67		
		R_{50}		0.65	0.58	0.52	0.47	0.33		1.00	0.92	0.92	0.83	0.75		
AFR		R_1	53	0.19	0.19	0.15	0.15	0.08	8	0.50	0.50	0.50	0.50	0.25		
		R_{10}		0.51	0.45	0.36	0.28	0.26		0.88	0.88	0.63	0.50	0.25		
		R_{25}		0.60	0.55	0.51	0.43	0.32		0.88	0.88	0.88	0.88	0.50		
		R_{50}		0.70	0.58	0.55	0.51	0.43		0.88	0.88	0.88	0.88	0.75		
AFR males		R_1	25	0.12	0.12	0.08	0.08	0.08	5	0.20	0.20	0.20	0.20	0.20		
		R_{10}		0.56	0.56	0.36	0.24	0.20		0.80	0.80	0.40	0.20	0.20		
		R_{25}		0.60	0.56	0.56	0.52	0.32		0.80	0.80	0.80	0.80	0.60		
		R_{50}		0.76	0.60	0.56	0.56	0.52		0.80	0.80	0.80	0.80	0.80		
AFR females		R_1	28	0.25	0.25	0.21	0.21	0.07	3	1.00	1.00	1.00	1.00	0.33		
		R_{10}		0.46	0.36	0.36	0.32	0.32		1.00	1.00	1.00	1.00	1.00		
		R_{25}		0.61	0.54	0.46	0.36	0.32		1.00	1.00	1.00	1.00	1.00		
		R_{50}		0.64	0.57	0.54	0.46	0.36		1.00	1.00	1.00	1.00	1.00		
CAU		R_1	25	0.64	0.64	0.52	0.52	0.40	25	0.64	0.64	0.52	0.52	0.40		
		R_{10}		0.80	0.72	0.72	0.64	0.56		0.80	0.72	0.72	0.64	0.56		
		R_{25}		0.88	0.84	0.80	0.76	0.60		0.88	0.84	0.80	0.76	0.60		
		R_{50}		0.92	0.88	0.88	0.80	0.76		0.92	0.88	0.88	0.80	0.76		
CAU males		R_1	16	0.63	0.63	0.50	0.50	0.38	16	0.63	0.63	0.50	0.50	0.38		
		R_{10}		0.81	0.75	0.75	0.63	0.56		0.81	0.75	0.75	0.63	0.56		
		R_{25}		0.88	0.81	0.81	0.81	0.63		0.88	0.81	0.81	0.81	0.63		
		R_{50}		0.88	0.88	0.88	0.81	0.81		0.88	0.88	0.88	0.81	0.81		
CAU females		R_1	9	0.67	0.67	0.56	0.56	0.44	9	0.67	0.67	0.56	0.56	0.44		
		R_{10}		0.78	0.67	0.67	0.67	0.56		0.78	0.67	0.67	0.67	0.56		
		R_{25}		0.89	0.89	0.78	0.67	0.56		0.89	0.89	0.78	0.67	0.56		
		R_{50}		1.00	0.89	0.89	0.78	0.67		1.00	0.89	0.89	0.67	0.56		
HIS		R_1	50	0.18	0.16	0.14	0.14	0.08	0	—	—	—	—	—		
		R_{10}		0.36	0.30	0.24	0.22	0.20		—	—	—	—	—		
		R_{25}		0.44	0.38	0.34	0.26	0.22		—	—	—	—	—		
		R_{50}		0.52	0.46	0.40	0.36	0.26		—	—	—	—	—		
HIS males		R_1	27	0.22	0.19	0.15	0.15	0.07	0	—	—	—	—	—		
		R_{10}		0.41	0.37	0.30	0.30	0.26		—	—	—	—	—		

Table 2 (continued)

		SP3.5 and SP6.0					SP3.5					
R_k	t	($g = 6159$)	($g = 12,000$)	($g = 25,000$)	($g = 50,000$)	($g = 108,000$)	t	($g = 6159$)	($g = 12,000$)	($g = 25,000$)	($g = 50,000$)	($g = 108,000$)
HIS females	R_{25}	0.44	0.41	0.37	0.33	0.30		—	—	—	—	—
	R_{50}	0.52	0.44	0.44	0.37	0.33		—	—	—	—	—
	R_1	0.13	0.13	0.13	0.13	0.09	0	—	—	—	—	—
	R_{10}	0.30	0.22	0.17	0.13	0.13		—	—	—	—	—
	R_{25}	0.43	0.35	0.30	0.17	0.13		—	—	—	—	—
18–29 Years	R_{50}	0.52	0.48	0.35	0.35	0.17		—	—	—	—	—
	R_1	0.37	0.36	0.29	0.29	0.17	29	0.59	0.59	0.48	0.48	0.31
	R_{10}	0.58	0.51	0.46	0.41	0.37		0.79	0.72	0.66	0.59	0.52
	R_{25}	0.67	0.62	0.56	0.51	0.42		0.86	0.83	0.79	0.76	0.62
	R_{50}	0.72	0.68	0.64	0.58	0.51		0.90	0.86	0.86	0.79	0.76
18–29 Years males	R_1	0.39	0.37	0.28	0.28	0.20	19	0.53	0.53	0.42	0.42	0.32
	R_{10}	0.63	0.59	0.52	0.46	0.41		0.79	0.74	0.63	0.53	0.47
	R_{25}	0.67	0.63	0.61	0.61	0.50		0.84	0.79	0.79	0.79	0.63
	R_{50}	0.72	0.67	0.67	0.61	0.61		0.84	0.84	0.84	0.79	0.79
	R_1	0.34	0.34	0.31	0.31	0.13	10	0.70	0.70	0.60	0.60	0.30
18–29 Years females	R_{10}	0.50	0.41	0.38	0.34	0.31		0.80	0.70	0.70	0.70	0.60
	R_{25}	0.66	0.59	0.50	0.38	0.31		0.90	0.90	0.80	0.70	0.60
	R_{50}	0.72	0.69	0.59	0.53	0.38		1.00	0.90	0.90	0.80	0.70
	R_1	0.13	0.13	0.10	0.10	0.10	2	1.00	1.00	1.00	1.00	1.00
	R_{10}	0.40	0.33	0.25	0.20	0.18		1.00	1.00	1.00	1.00	1.00
30–49 Years	R_{25}	0.48	0.43	0.40	0.28	0.20		1.00	1.00	1.00	1.00	1.00
	R_{50}	0.63	0.45	0.43	0.40	0.28		1.00	1.00	1.00	1.00	1.00
	R_1	0.06	0.06	0.06	0.06	0.06	1	1.00	1.00	1.00	1.00	1.00
	R_{10}	0.39	0.39	0.22	0.17	0.11		1.00	1.00	1.00	1.00	1.00
	R_{25}	0.44	0.39	0.39	0.28	0.17		1.00	1.00	1.00	1.00	1.00
30–49 Years males	R_{50}	0.67	0.44	0.39	0.39	0.28		1.00	1.00	1.00	1.00	1.00
	R_1	0.18	0.18	0.14	0.14	0.14	1	1.00	1.00	1.00	1.00	1.00
	R_{10}	0.41	0.27	0.27	0.23	0.23		1.00	1.00	1.00	1.00	1.00
	R_{25}	0.50	0.45	0.41	0.27	0.23		1.00	1.00	1.00	1.00	1.00
	R_{50}	0.59	0.45	0.45	0.41	0.27		1.00	1.00	1.00	1.00	1.00
≥ 50 Years	R_1	0.10	0.10	0.10	0.10	0.10	2	0.50	0.50	0.50	0.50	0.50
	R_{10}	0.40	0.40	0.30	0.20	0.20		1.00	1.00	1.00	1.00	1.00
	R_{25}	0.50	0.40	0.40	0.40	0.20		1.00	1.00	1.00	1.00	1.00
	R_{50}	0.50	0.50	0.40	0.40	0.40		1.00	1.00	1.00	1.00	1.00
	R_1	0.00	0.00	0.00	0.00	0.00	1	0.00	0.00	0.00	0.00	0.00
≥ 50 Years males	R_{10}	0.50	0.50	0.25	0.00	0.00		1.00	1.00	1.00	1.00	1.00
	R_{25}	0.50	0.50	0.50	0.50	0.00		1.00	1.00	1.00	1.00	1.00
	R_{50}	0.50	0.50	0.50	0.50	0.50		1.00	1.00	1.00	1.00	1.00
	R_1	0.50	0.50	0.50	0.50	0.50		1.00	1.00	1.00	1.00	1.00
	R_{10}	0.50	0.50	0.50	0.50	0.50		1.00	1.00	1.00	1.00	1.00

Table 2 (continued)

SP3.5 and SP6.0											
SP3.5											
R_k	t	$(g = 6159)$	$(g = 12,000)$	$(g = 50,000)$	$(g = 108,000)$	t	$(g = 6159)$	$(g = 12,000)$	$(g = 50,000)$	$(g = 108,000)$	
≥50 Years females	R_1	6	0.17	0.17	0.17	1	1.00	1.00	1.00	1.00	
	R_{10}		0.33	0.33	0.33		1.00	1.00	1.00	1.00	
	R_{25}		0.50	0.33	0.33		1.00	1.00	1.00	1.00	
	R_{50}		0.50	0.33	0.33		1.00	1.00	1.00	1.00	
SP6.0											
	R_k	t	$(g = 6159)$				$(g = 50,000)$				$(g = 108,000)$
Overall	R_1	95	0.16	0.15	0.12	0.12	0.12	0.12	0.12	0.06	
	R_{10}		0.40	0.34	0.27	0.23	0.27	0.23	0.23	0.21	
	R_{25}		0.49	0.43	0.39	0.31	0.39	0.31	0.31	0.23	
	R_{50}		0.59	0.49	0.44	0.40	0.44	0.40	0.40	0.31	
Cohorts											
Male	R_1	47	0.17	0.15	0.11	0.11	0.11	0.11	0.11	0.06	
	R_{10}		0.45	0.43	0.32	0.28	0.32	0.28	0.28	0.23	
	R_{25}		0.49	0.45	0.43	0.38	0.43	0.38	0.38	0.28	
	R_{50}		0.62	0.49	0.47	0.43	0.47	0.43	0.43	0.38	
Female	R_1	48	0.15	0.15	0.13	0.13	0.13	0.13	0.13	0.06	
	R_{10}		0.35	0.25	0.23	0.19	0.23	0.19	0.19	0.19	
	R_{25}		0.50	0.42	0.35	0.23	0.35	0.23	0.23	0.19	
	R_{50}		0.56	0.50	0.42	0.38	0.42	0.38	0.38	0.23	
AFR	R_1	45	0.13	0.13	0.09	0.09	0.09	0.09	0.09	0.04	
	R_{10}		0.44	0.38	0.31	0.24	0.31	0.24	0.24	0.22	
	R_{25}		0.56	0.49	0.44	0.36	0.44	0.36	0.36	0.24	
	R_{50}		0.67	0.53	0.49	0.44	0.49	0.44	0.44	0.36	
AFR males	R_1	20	0.10	0.10	0.05	0.05	0.05	0.05	0.05	0.05	
	R_{10}		0.50	0.50	0.35	0.25	0.35	0.25	0.25	0.20	
	R_{25}		0.55	0.50	0.50	0.45	0.50	0.45	0.45	0.25	
	R_{50}		0.75	0.55	0.50	0.50	0.50	0.50	0.50	0.45	
AFR females	R_1	25	0.16	0.16	0.12	0.12	0.12	0.12	0.12	0.04	
	R_{10}		0.40	0.28	0.28	0.24	0.28	0.24	0.24	0.24	
	R_{25}		0.56	0.48	0.40	0.28	0.40	0.28	0.28	0.24	
	R_{50}		0.60	0.52	0.48	0.40	0.40	0.28	0.28	0.24	
CAU	R_1	0	—	—	—	—	—	—	—	—	

Table 2 (continued)

R_k	t	SP6.0				
		($g = 6159$)	($g = 12,000$)	($g = 25,000$)	($g = 50,000$)	($g = 108,000$)
CAU males	0	R_{10}	—	—	—	—
		R_{25}	—	—	—	—
		R_{50}	—	—	—	—
		R_I	—	—	—	—
		R_{10}	—	—	—	—
CAU females	0	R_{25}	—	—	—	—
		R_{50}	—	—	—	—
		R_1	—	—	—	—
		R_{10}	—	—	—	—
		R_{25}	—	—	—	—
HIS	50	R_{50}	—	—	—	—
		R_1	0.18	0.16	0.14	0.14
		R_{10}	0.36	0.30	0.24	0.22
		R_{25}	0.44	0.38	0.34	0.26
		R_{50}	0.52	0.46	0.40	0.36
HIS males	27	R_1	0.22	0.19	0.15	0.15
		R_{10}	0.41	0.37	0.30	0.30
		R_{25}	0.44	0.41	0.37	0.33
		R_{50}	0.52	0.44	0.44	0.37
		R_1	0.13	0.13	0.13	0.13
HIS females	23	R_{10}	0.30	0.22	0.17	0.13
		R_{25}	0.43	0.35	0.30	0.17
		R_{50}	0.52	0.48	0.35	0.35
		R_1	0.24	0.22	0.18	0.18
		R_{10}	0.45	0.39	0.35	0.31
18–29 Years	49	R_{25}	0.55	0.49	0.43	0.37
		R_{50}	0.61	0.57	0.51	0.45
		R_1	0.30	0.26	0.19	0.19
		R_{10}	0.52	0.48	0.44	0.41
		R_{25}	0.56	0.52	0.48	0.48
18–29 Years males	27	R_{50}	0.63	0.56	0.56	0.48
		R_1	0.18	0.18	0.18	0.18
		R_{10}	0.36	0.27	0.23	0.18
		R_{25}	0.55	0.45	0.36	0.23
						0.18
18–29 Years females	22					

Table 2 (continued)

		SP6.0						
		R_k	t	($g = 6159$)	($g = 12,000$)	($g = 25,000$)	($g = 50,000$)	($g = 108,000$)
30–49 Years	R_{50}			0.59	0.59	0.45	0.41	0.23
	R_1		38	0.08	0.08	0.05	0.05	0.05
	R_{10}			0.37	0.29	0.21	0.16	0.13
	R_{25}			0.45	0.39	0.37	0.24	0.16
	R_{50}			0.61	0.42	0.39	0.37	0.24
30–49 Years males	R_1		17	0.00	0.00	0.00	0.00	0.00
	R_{10}			0.35	0.35	0.18	0.12	0.06
	R_{25}			0.41	0.35	0.35	0.24	0.12
	R_{50}			0.65	0.41	0.35	0.35	0.24
	R_1		21	0.14	0.14	0.10	0.10	0.10
30–49 Years females	R_{10}			0.38	0.24	0.24	0.19	0.19
	R_{25}			0.48	0.43	0.38	0.24	0.19
	R_{50}			0.57	0.43	0.43	0.38	0.24
	R_1		8	0.00	0.00	0.00	0.00	0.00
	R_{10}			0.25	0.25	0.13	0.13	0.13
≥ 50 Years	R_{25}			0.38	0.25	0.25	0.25	0.13
	R_{50}			0.38	0.38	0.25	0.25	0.25
	R_1		3	0.00	0.00	0.00	0.00	0.00
	R_{10}			0.33	0.33	0.00	0.00	0.00
	R_{25}			0.33	0.33	0.33	0.33	0.00
≥ 50 Years males	R_{50}			0.33	0.33	0.33	0.33	0.33
	R_1		5	0.00	0.00	0.00	0.00	0.00
	R_{10}			0.20	0.20	0.20	0.20	0.20
	R_{25}			0.40	0.20	0.20	0.20	0.20
	R_{50}			0.40	0.40	0.20	0.20	0.20
total test subjects								

(8–24 %). In contrast, the 30–49- (5–8 %) and ≥ 50 -year-old (0 %) cohorts exhibited moderately lower R_1 identification rates, although, again, not significantly different (all $p > 0.28$) than the overall rates. Notable is that, depending on the gallery size employed, in 31–59 % of the cases, a correct life photo match for a CT-derived facial image was located within the top 50 images (R_{50}) of the generated candidate lists for this protocol. Examination of cohort relationships for SP6.0, indicated that few significant performance differences (majority of $p > 0.12$) exist between the sexes or across the ancestral and age cohorts eligible (i.e., $n \geq 5$) for examination, irrespective of R_k , gallery size, or cohort sub-discrimination.

Performance between Scan Protocols

Examination of the overall performance (Table 2) between protocols SP3.5 and SP6.0 indicated that highly significant differences (all $p < 0.01$) in identification and candidate list inclusion rates exist *between* the protocols at all levels of R_k and gallery size. Examination of cohort relationships indicated that, with the exception of the 18–29-year-old male cohort (all $p > 0.06$), highly significant performance differences (majority of $p < 0.01$) were also observed between protocols across most R_k and gallery size for all demographic cohorts eligible for examination (i.e., $n \geq 5$).

Performance by Gallery Size and Rank Class Pairs

A series of McNemar's tests conducted to investigate the relationship between gallery size and rank class revealed that highly significantly different (majority of $p \leq 0.01$) overall (i.e., rates for SP3.5 and SP6.0 combined) identification and candidate list inclusion rates exist across most gallery and rank class pairs. When performance for each scanning protocol was examined separately, a general pattern of non-significance (majority of $p \geq 0.06$) was observed across most gallery and rank class pairs for SP3.5. In contrast, an overall pattern of highly significant performance differences (majority of $p \leq 0.01$) was observed across gallery and rank pairs for SP6.0. Noteworthy is that, in general, few significant differences in identification rates were observed across gallery pairs at the R_1 rank class within either scanning protocol.

Conclusion

The recognizability of facial images extracted from publically available medical scans raises concerns for the protection of an individual's personally identifiable information. The results of this research suggest that the probability of matching a CT-derived facial image with a corresponding life photo is moderately high, even when blind searching large photo pools.

Several observations are offered. First, depending on the scan protocol and gallery size employed, in 31–91 % of the cases, a correct life photo match for a CT-derived facial image was located within the top 50 images (R_{50}) of the generated candidate lists, even when searching a gallery in excess of 100,000 photos. Second, highly significant differences in identification rates were observed between the two scanning protocols examined, with the facial images extracted from SP6.0 resulting in candidate list inclusion rates significantly lower than the rates observed for SP3.5. Initially, these results appear to suggest that protocol SP6.0 confers a degree of protection against identification (i.e., lower candidate list inclusion rates) and, as such, would be a preferred imaging protocol, if applicable to a given medical condition for which the scan is required. However, the sample sizes and demographic distribution disparities between the scanning protocols employed preclude inference regarding which protocol, if any, has the potential to provide greater protection against automated identification. Further examination is needed to more fully assess the impact of this variable. Finally, the results suggest that the facial image data inherent in commonly employed medical imaging modalities may need consideration as a potentially identifiable form of comparable facial imagery protected under privacy statutes such as those provided in the HIPAA legislation. This consideration is likely more critical than ever given the growing availability of facial recognition services freely provided by popular social media and photo sharing sites that host hundreds of billions of photos each year [37–39], millions of which contain potentially harvestable facial images.

Acknowledgements This is publication 16-19 of the Federal Bureau of Investigation (FBI). The authors had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. This research was supported in part by an appointment to the Visiting Scientist Program at the FBI Laboratory Division, administered by the Oak Ridge Institute for Science and Education, through an interagency agreement between the US Department of Energy and the FBI. Names of commercial manufacturers are provided for identification purposes only and inclusion does not imply endorsement by the FBI. The views expressed are those of the authors and do not necessarily reflect the official policy or position of the FBI or the US Government. We thank reviewers Richard M. Thomas and Leslie L. Cavis (both FBI) for their helpful comments. We also thank Lisa G. Bailey (FBI) for providing the composites used as example facial images in this manuscript, the Pinellas County Sheriff's Office (Florida) for use of their facial image database, and the NEC engineering and support staff for the loan of the NeoFace® Reveal system and support of this research.

References

1. Smith-Bindman R, Miglioretti DL, Johnson E, Lee C, Feigelson HS, Flynn M, Greenlee RT, Kruger RL, Hombrook MC, Roblin D, Solberg LI, Vanneman N, Weinmann S, Williams AE: Use of diagnostic imaging studies and associated radiation exposure for

- patients enrolled in large integrated health care systems, 1996–2010. *JAMA* 307(22):2400–2409. 2012
2. Rao VM, Levin DC, Parker L, Frangos AJ, Sunshine JH: Trends in utilization rates of the various imaging modalities in emergency departments: Nationwide Medicare data from 2000 to 2008. *J Am Coll Radiol* 8(10):706–709. 2011
3. Agarwal R, Bergey M, Sonnad S, Butowsky H, Bhargavan M, Bleshtman MH: Inpatient CT and MRI utilization: trends in the academic hospital setting. *J Am Coll Radiol* 7(12):949–955. 2010
4. Broder J, Warshawer DM: Increasing utilization of computed tomography in the adult emergency department, 2000–2005. *Emerg Radio* 13(1):25–30. 2006
5. Brenner DJ, Hall EJ: Computed tomography—an increasing source of radiation exposure. *N Engl J Med* 357(22):2277–2284. 2007
6. The Cancer Imaging Archive. Available at <http://www.cancerimagingarchive.net>. Accessed January 2016
7. Peddicord D, Waldo AB, Boutin M, Grande T, Gutierrez L Jr: A proposal to protect privacy of health information while accelerating comparative effectiveness research. *Health Aff (Millwood)* 29(11):2082–2090. 2010
8. Budin F, Zeng D, Ghosh A, Bullitt E: Preventing facial recognition when rendering MR images of the head in three dimensions. *Med Image Anal* 12(3):229–239. 2008
9. Noumeir R, Lemay A, Lina JM: Pseudonymization of radiology data for research purposes. *J Digit Imaging* 20(3):284–295. 2007
10. P.L. No. 104–191, 110 Stat. 1938 (1996). United States Department of Health and Human Services. The Health Insurance Portability and Accountability Act (HIPAA)
11. 11. C.F.R. § 164.514(e) (2). US Department of Health and Human Services. Privacy of Individually Identifiable Health Information. Other requirements relating to uses and disclosures of protected health information
12. Fetzer DT, West OC: The HIPAA privacy rule and protected health information: implications in research involving DICOM image databases. *Acad Radiol* 15(3):390–395. 2008
13. Avrin D: HIPAA privacy and DICOM anonymization for research. *Acad Radiol* 15(3):273. 2008
14. Maal TJJ, Plooi JM, Rangel, FA Mollemans W, Schutyser FAC, Bergé SJ: The accuracy of matching three-dimensional photographs with skin surfaces derived from cone-beam computed tomography. *Int J Oral Maxillofac Surg* 37(7):641–646. 2008
15. Jen-Sho Chen J, Juluru K, Morgan T, Moffitt R, Siddiqui KM, Siegel EL: Implications of surface-rendered facial CT images in patient privacy. *Am J Roentgenol* 202(6):1267–1271. 2014
16. OsiriX . DICOM Viewer for Mac. Available at <http://www.osirix-viewer.com/Downloads.html>. Accessed August 2016
17. MicroDicom. DICOM Viewer for Windows. Available at <http://www.microdicom.com>. Accessed August 2016
18. Gingo CADx. Cross Platform DICOM Viewer. Available at <http://ginkgo-cadx.com/en>. Accessed August 2016
19. Prior FW, Brunson B, Hildebolt C, Nolan TS, Pringle M, Vaishnavi SN, Larson-Prior LJ: Facial recognition from volume-rendered magnetic resonance imaging data. *IEEE Trans Inf Technol Biomed* 13(1):5–9. 2009
20. Mazura JC, Julura K, Chen JJ, Morgan TA, John M, Siegel EL: Facial recognition software success rates for the identification of 3D surface reconstructed facial images: implications for patient privacy and security. *J Digit Imaging* 25(3):347–351. 2012
21. Hwang HG, Han HE, Kuo KM, Liu CF: The differing privacy concerns regarding exchanging electronic medical records of internet users in Taiwan. *J Med Syst* 36(6):3783–93. 2012
22. Senior A, Pankanti S: Privacy protection and face recognition. In: Li, SZ and Jain, AK, eds: *Handbook of face recognition*, 2nd ed, NY: Springer, pp 671–691. 2011
23. American Civil Liberties Union. Washington DC. Available at <https://www.aclu.org>. Accessed August 2016
24. Electronic Privacy Information Center. Washington DC. Available at <https://epic.org>. Accessed August 2016
25. Electronic Frontier Foundation. San Francisco, California. Available at <https://www.eff.org>. Accessed August 2016
26. Parks CP, Richard H, Monson KL: Preliminary performance assessment of computer automated facial approximations using computed tomography scans of living individuals. *Forensic Sci Int* 233(1–3):133–139. 2013
27. Picasa: Available at <http://picasa.en.softonic.com>. Accessed August 2016
28. NeoFace® Reveal: Available at <https://www.necam.com/Biometrics/doc.cfm?t=NeoFaceReveal>. Accessed August 2016
29. Parks CL, Monson KL: Recognizability of ReFace facial approximations in an automated facial recognition context for potential application in unidentified persons data repositories. Part I: Optimally modeled conditions. *Forensic Sci Int* (in review)
30. Parks CL, Monson KL: Recognizability of ReFace facial approximations in an automated facial recognition context for potential application in unidentified persons data repositories. Part II: Operationally modeled conditions. *Forensic Sci Int* (in review)
31. P.J. Phillips, P. Grother, R. Michaels: Evaluation methods in face recognition. In: Li, SZ and Jain, AK, eds: *Handbook of face recognition*, 2nd ed, NY: Springer, pp 551–574. 2011.
32. Jain, AK and Ross, AA: Introduction to biometrics. In: Jain, AK, Flynn, P and Ross, AA, eds: *Handbook of biometrics*, NY: Springer, pp 1–22. 2008
33. Phillips PJ, Grother P, Micheals RJ, Blackburn DM, Tabassi E, Bone JM: Face Recognition Vendor Test (FRVT). Evaluation Report. National Institute of Standards and Technology Interagency Report 6965. 2002 Available at http://www.nist.gov/customcf/get_pdf.cfm?pub_id=50767. Accessed August 2016
34. Grother PJ, Quinn GW, Phillips PJ: Report on the evaluation of 2D still-image face recognition algorithms. Technical Report. National Institute of Standards and Technology Interagency Report 7709. Available at http://biometrics.nist.gov/cs_links/face/mbe/MBE_2_D_face_report_NISTIR_7709.pdf. Accessed August 2016
35. Grother PJ, Ngan ML: Face Recognition Vendor Test (FRVT). Performance of Face Identification Algorithms. National Institute of Standards and Technology Interagency Report 8009. Available at http://www.nist.gov/customcf/get_pdf.cfm?pub_id=915761. Accessed August 2016
36. Holm S: A simple sequentially rejective multiple test procedure. *Scand J Statistics* 6(2):65–70. 1979
37. Ho K: 41 Up-to-date Facebook facts and stats. Available at <http://blog.wishpond.com/post/115675435109/40-up-to-date-facebook-facts-and-stats>. Accessed August 2016
38. Robertson A: Facebook users have uploaded a quarter-trillion photos since the site's launch. Available at <http://www.theverge.com/2013/9/17/4741332/facebook-users-have-uploaded-a-quarter-trillion-photos-since-launch>. Accessed August 2016
39. Kotenko J: Facebook reveals we upload a whopping 350 million photos to the network daily. Available at <http://www.digitaltrends.com/social-media/according-to-facebook-there-are-350-million-photos-uploaded-on-the-social-network-daily-and-thats-just-crazy>. Accessed August 2016