Interoperability between Fingerprint Biometric Systems: An Empirical Study

Stephen Mason, Ilir Gashi
Centre for Software Reliability
City University London
London, United Kingdom
{stephen.mason.1, ilir.gashi.1}@city.ac.uk

Luca Lugini, Emanuela Marasco, Bojan Cukic
Lane Department of
Computer Science and Electrical Engineering
West Virginia University
Morgantown, WV (USA)
{emanuela.marasco, bojan.cukic}@mail.wvu.edu
lulugini@mix.wvu.edu

***Abstract*—Fingerprints are likely the most widely used biometric in commercial as well as law enforcement applications. With the expected rapid growth of fingerprint authentication in mobile devices their importance justifies increased demands for dependability. An increasing number of new sensors, applications and a diverse user population also intensify concerns about the interoperability in fingerprint authentication. In most applications, fingerprints captured for user enrollment with one device may need to be "matched" with fingerprints captured with another device. We have performed a large-scale study with 494 participants whose fingerprints were captured with 4 different industry-standard optical fingerprint devices. We used two different image quality algorithms to evaluate fingerprint images, and then used three different matching algorithms to calculate match scores. In this paper we present a comprehensive analysis of dependability and interoperability attributes of fingerprint authentication and make empirically-supported recommendations on their deployment strategies.**

Keywords - biometric systems; empirical assessment; experimental results; design diversity; interoperability

I. INTRODUCTION

Fingerprint-based user authentication is one of the most prolific commercial branches of biometrics. Since the authentication process needs two samples from each user, most systems need to anticipate that the device used for a user's *enrollment* (creation of the so called *gallery image* or template) may not be the same as the device used at the time of *identification* or identity *verification* (so called *probe image* or template). Fingerprints can be acquired through different live-scan sensing technologies: *optical*, *solid-state* and *ultrasound* [1]. In optical sensors, the finger is placed on the surface of a transparent prism which is typically illuminated through the left side and the image is taken through a camera. The light entering the prism is reflected at the *valleys* and absorbed at the *ridges* of a fingerprint. In *solid-state* devices, the finger is modeled as the upper electrode of the capacitor, while the metal plate is modeled as the lower electrode. The variation in capacity between valleys and ridges can be measured when the finger is placed on the sensor. In the case of swipe *solid-state* sensors, impressions are obtained by swiping the finger on the surface of the sensor. *Ultrasound* sensors exploit the difference of acoustic impedance between the skin of the ridges and the air in the valleys of a finger.

Even within the specific sensing technology, the acquisition quality may vary across sensors [5]. Different arrangements of sensing elements in each device may introduce variations and distortions in the biometric data. In particular, differences in resolution and scanning area impact the *feature set* [1] extracted from the acquired image. A biometric matching system is required to handle variations introduced in the biometric data through different devices [1]. While device diversity is to be expected, commercial fingerprint matchers typically show a decrease in inter-device matching performance [3]. For systems deployed at US international airports, sensor interoperability is important. In this application, fingerprints are currently enrolled using a 500 dpi optical sensor with a sensing area of 1.2" x 1.2". As different devices may be used for verification, limited interoperability between the devices is a significant concern. Interoperability grows in importance as the scale of adoption of biometric devices and the pace of innovation increase: older biometric devices get replaced with newer designs, but the samples enrolled with older devices remain in operational use.

In this paper we provide a comprehensive analysis of the effects of interoperability on the overall dependability of the fingerprint matchers. Failures in our case are defined as *false-matches* (a fingerprint is judged to belong to a person when in fact this is not the case) and *false-non-matches* (a fingerprint is judged to not belong to a person when in fact this is the case). We use data from a large-scale study in which we collected all ten fingerprints of 494 participants using 4 different biometric devices. The analysis in this paper uses right point fingers only. The use of one finger is typical for authentication applications. The sample of 494 is large because we are dealing with human participants and we follow a properly approved collection protocol that requires volunteers to dedicate one hour of time to biometric data collection for which they are adequately compensated. The fingerprints were captured twice per person: once for the purpose of creating the *enrollment* or *gallery* image and the

[1] A feature extractor module extracts the information from the fingerprint image by detecting characteristics representative of the identity with respect to the matching process. These characteristics are referred to as a *feature set* and are used by a pattern classifier to make the decision about the identity of the user.

second time for the purpose of creating the *probe* image for identification or authentication. The quality of each image is evaluated using two different image quality algorithms. In the analysis, we *match* the probe and gallery images. We use three different off-the-shelf products to calculate the match scores for each pair of images. This gave us a rich dataset for interoperability analysis, which allowed us to categorize and rank results: per soft-biometric (gender and age); per fingerprint capture device; per matcher and per fingerprint quality evaluation algorithm. We then performed exploratory analysis to identify various cause-and-effect patterns in our data set. This enabled us to make empirically-supported recommendations on optimal deployments for different scenarios, depending on the flexibility and choice (in terms of fingerprint capture devices, fingerprint quality algorithms and matching algorithms) for a given application.

The rest of the paper is organized as follows: Section II presents background and related work; Section III describes the experimental setup; Section IV presents the results of our analysis; Section V presents a discussion of the empirically-supported deployment recommendations we can make based on our observations. Finally Section VI presents conclusions and provisions for further work.

II. Background and Related Work

Recent investigations focused on the impact of diverse fingerprints capture platforms and soft-biometrics on match error rates.

1) *Interoperability Literature*

In [2] the authors statistically measured the degree of change in match scores across different optical devices. Results of the Kendall's rank correlation test pointed out that there is a significant difference between sensor pairs and that the change is not symmetric when inverting the two devices. In [3] the authors proposed a learning-based approach to improve cross-device fingerprint verification performance. They extracted quality and intensity-based characteristics of fingerprint images acquired using four different commercial optical devices and scanned ink rolled prints. They were concatenated with the match score into a feature vector used for training a pattern classifier. The model was developed for both intra-device and cross-device matching for all device pairs. Poh *et al.* designed a Bayesian Belief Network (BBN) to estimate the posterior probability of the device d given quality q, referred to as $p(d|q)$ [4]. Clustering is applied to each device to explain hidden quality factors. However, such data clustering does not explicitly model the influence of each device. Jain and Ross considered the interoperability issue as one related to the variability induced in the feature set when different sensor technologies are used (e.g. optical vs. capacitive) [5]. When matching images acquired by Digital Biometrics and Veridicom sensors, they reported an Equal Error Rate (EER) of 23.13%, compared to an EER of 6.14% and 10.39% when using only Digital Biometrics and Veridicom, respectively. Ross and Nadgir subsequently proposed a compensation model which computes the relative distortion between images acquired using different devices [6]. The model is based on a thin-plate spline whose parameters rely on control points manually selected in order to cover representative areas where distortions can occur in the fingerprint image. Their method is, therefore, not completely automated.

2) *Age/Gender Literature.*

Past studies examined effects of fingerprints from different age groups and gender [7]. Effects of ageing impact the quality of fingerprints. Over the life of the individual, the skin becomes drier and thinner; reduction of collagen causes skin wilting. These factors affect the sample provided to the fingerprint sensor [8]. Age affects the differences in quality of the physical state of the fingerprint (e.g., skin deterioration), while the ridge/valley pattern is believed to remain stable over the life time of an individual. Regarding gender, most of the works analyzed ridges in the spatial domain. They observed that females present a higher ridge density compared to males, due to finer epidermal ridge details. Ridge density is defined as the number of ridges which occurs in a certain space [9]. In 1999, Acree manually counted ridges in a well-defined area [10].

In [11] [12] fingerprints are classified based on gender / age using statistics such as white lines count and ridge count that are manually extracted as proposed by Acree. Recently, a method based on both discrete wavelet transform (DWT) and singular value decomposition (SVD) has been proposed for gender and age estimation in [13].

3) *Biometric Fusion, parallel vs. sequential Literature*

The key to create a secure multibiometric system is in the design of the fusion scheme. The consolidation of biometric information can be performed at various levels: sensor level, feature extraction level, match score level, rank level (identification operation) and decision level [14][15][16][17][18]. Fusion may be able to provide better recognition results, if designed well. Combining match scores from different matchers has proved to be an effective fusion strategy because it offers the trade-off between the information content and the ease of fusion implementation. The benefits of multibiometrics depend on the diversity of component information [19][20]. Intuitively, the classifier working with better data would produce better results than a classifier operating on noisy data. Hence, researchers introduced quality-based fusion schemes, where the quality measures of the samples are incorporated in the fusion to improve performance [21].

An interesting open research issue pertains to the estimation of the decision reliability with respect to the fusion scheme [22]. The sequential fusion strategy considers systems sequentially, and the goal is to make decisions by employing as few systems as possible. Research suggests that participating systems should be ordered by decreasing match confidence [24] [25]. Better understanding of differences between fingerprint systems studied in this paper may allow us to develop effective multi-sensor fusion schemes.

III. Experimental Setup

The dataset we use has been collected in 2012. Each participant's fingerprints were captured by multiple optical sensors. The order of use of fingerprint scanners was the same for all 494 volunteers. Each of them self-reported gender and age.

Figure 1 shows the number of participants (x-axis) by age (y-axis) and gender (blue=male; red=female). The gender ratio is almost equal, with 52% male participants. The majority of participants in the dataset are young, with 50% aged 27 and below.

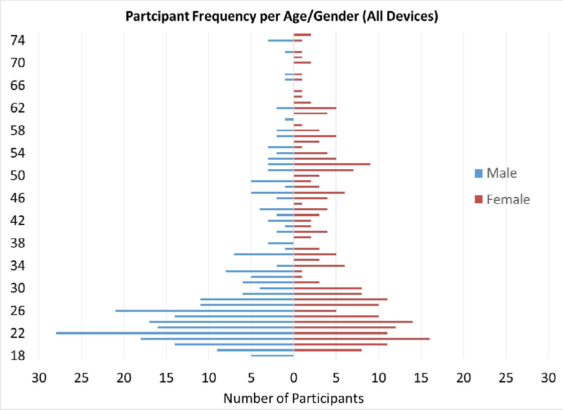


Figure 1. Christmas tree diagram showing the number of participants (x-axis) for each age (y-axis) and gender (bar colour) in the dataset.

Fingerprints were acquired using four Live-scan devices (D0 – D3[2], see Table I). The choice of these devices was partially influenced by availability (we have them in our lab) and sponsor's interest, but we should stress that all of them are high-end devices, widely used in industry and hence representative of common real world installations. For each Live-scan device participants provided two *sets* of fingerprints (in the same lab visit, i.e. one after the other), for each device consisting of: rolled individual fingers on both hands, left slap (i.e. slapping the four (non-thumb) fingers on the device), right slap, and thumbs slap. The optical sensor utilizes a glass platen, a laser light-source and a Charge-Coupled Device (CCD) or a Complementary Metal–Oxide–Semiconductor (CMOS) camera for constructing fingerprint images. When finger is placed on the glass platen, a laser light is reflected through the prism and facilitates the imaging. Fingerprints were collected without controlling the quality or the position of the finger. For the purpose of the analysis in this paper we have used the right hand's index fingerprints only.

Fingerprint image quality was assessed using two different quality algorithms:

- The NIST (National Institute for Standards and Technology) Fingerprint Image Quality (NFIQ) algorithm[3] [23]: an open source tool that has become the industry standard for fingerprint image quality assessment; the quality is classified into five levels, 1 (highest) to 5 (lowest).
- The MITRE IQF [27] uses the two-dimensional frequency information of the image, the power spectrum, to assess fingerprint quality. The quality score ranges between 0 (lowest) and 100 (highest).

While the NFIQ quality score predicts the impact that the image has on the matching system performance in terms of error rates, the MITRE IQF score gives an assessment of the visual quality of a fingerprint image. Figure 2 shows examples of the highest and lowest quality fingerprints, according to the NFIQ.

Table I. Characteristics of the live-scan devices used in our study.

	Device	Model	Resolution (dpi)	Image size (pixels)	Capture area (mm)
D0	Cross Match	Guardian R2	500	800 x 750	81 x 76
D1	i3	digID Mini	500	752 x 750	81 x 76
D2	L1 Identity Solutions	TouchPrint 5300	500	800 x 750	81 x 76
D3	Cross Match	Seek II	500	800 x 750	40.6 x 38.1

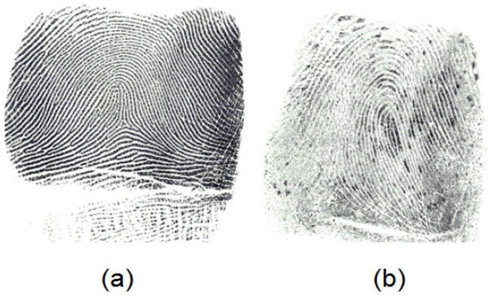


Figure 2. Fingerprint samples with NFIQ quality scores of (a) 1, best and (b) 5, worst.

We generated the match scores using three Commercial Off-the-Shelf (COTS) fingerprint matching products:

- (M1) Identix BioEngine Software Development Kit[4];
- (M2) Bozorth3, a minutiae based fingerprint matcher developed by NIST [23];
- (M3) BIO-key WEB-key Software Development Kit[5].

In the rest of the text, we will be using the abbreviations M1, M2 and M3 to refer to these three matching algorithms.

A matching algorithm compares two fingerprint images and returns a *score* based on the similarity between the two templates. The higher the score the more likely it is that the two templates come from the same finger.

The initial aims of our study are to compare the diversity that exists in the following dimensions:

[2] In the rest of the text, we will be using the abbreviations (D0-D3) to refer to these four respective devices. The same instance of each device were used throughout the data collection.

[3] http://www.nist.gov/itl/iad/ig/nbis.cfm

[4] http://www.morphotrust.com/pages/117-fingerprint-palm

[5] http://www.bio-key.com/products/overview-2/web-key

- Fingerprint capture devices
- Matching algorithms
- Image quality algorithms
- Age
- Gender

For each COTS matcher we are interested in the scores from two matching scenarios: *i)* comparing two fingerprints captured with the same device (intra-device), and *ii)* comparing two fingerprints captured with different devices (inter-device).

The notation reflecting the types of similarity match scores is shown in Table II. Since the total number of impostor scores could be very large, we limited it to a random subset which is still sufficient for statistical analysis. Table III reveals the number of scores in each category. We were unable to generate an equal number of impostor match scores (DMI/DDMI) for M3 because of license limitations, but the number is sufficient for the analysis.

Table II. Notation table for similarity score computations.

Device Match Genuine (DMG): Genuine match scores are generated when we match templates of the same user's right point finger. The image captured in the first user's interaction with a sensor is stored in the *gallery* (the database of fingerprint images which we search). The image acquired using the same device the second time is used as a *probe* (the set of images submitted for identification or verification). Since we have *494* participants and *4* devices, the total number of DMG scores is *1,976*.
Device Match Impostor (DMI): Impostor match scores are generated by matching the fingerprint template of a participant against those of all the other participants. DMI scores include only those in which both the gallery and probe images are acquired using the same device. The number of imposter scores is very large. We limit our analysis to randomly obtained *96,684* DMI match scores for matchers M1 and M2, and *77,616* DMI match scores for M3.
Diverse Device Match Genuine (DDMG): Genuine match scores generated when gallery and probe images are acquired using different devices. For each participant, having 4 collection sensors, we have 6 possible combinations with two match scores for each probe, resulting in a total of 5,928 match scores (494 * 2 * 6).
Diverse Device Match Impostor (DDMI): Impostor match scores generated using images from different devices.

Table III. Match score for different match scenarios.

Matching	**Participants**	**Number of devices**	**Samples**	**Similarity scores**		
				M1	**M2**	**M3**
DMG	*494*	4	*2*	*1,976*	*1,976*	*1,976*
DDMG	*494*	4	*2*	*5,928*	*5,928*	*5,928*
DMI	*494*	*4*	*2*	*96,684*	*96,684*	*77,616*
DDMI	*494*	*4*	*2*	*290,052*	*290,052*	*232,848*

IV. Exploratory Analysis

A. Image Quality Analysis

1) Soft-Biometrics (Age / Gender)

As explained previously, for each of the captured images in our dataset we obtained a fingerprint quality score from two different algorithms: NFIQ and MITRE. Our analysis of quality starts by investigating how the participants' soft-biometric traits, gender and age, affect the quality score. We have done an exhaustive analysis of fingerprint image quality per device, age and gender. Due to space constraints in the paper, we will only be able to summarize the main results. The complete analysis is provided in a technical report [26]. A reminder that the NFIQ algorithm assigns a score between 1 and 5 to each fingerprint image (1=highest quality; 5=lowest quality), and the MITRE algorithm assigns a score of 1-100 (1=lowest quality; 100=highest quality).

Looking at the NFIQ results, we found that on average the gender of the participant does affect the image quality assigned by NFIQ. Amongst the high NFIQ scores (1-2) the majority of images are from male participants; the lower quality scores (3-5) have a majority of images from female participants, as shown in Figure 3.

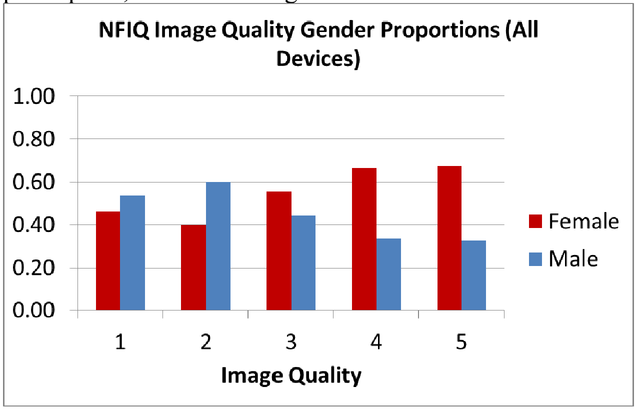


Figure 3. Bar chart showing the proportion (y-axis) of male (blue bars) and female (red bars) particpants for each NFIQ image quality score (x-axis) across all devices.

Figures 4 and 5 show the average NFIQ and MITRE quality scores respectively (x-axis) for all images by age (y-axis) and gender (blue=male; red=female) across all devices.

A general trend can be observed from these graphs, for both genders: the average image quality decreases as the participants' age increases[6]. However, we found that NFIQ is better determinant or discriminator of fingerprint image quality. In general, the variances in the quality scores assigned by the MITRE algorithm were relatively low, hence difficult to use it as a discriminator. In the image quality assessment process MITRE normalizes the power spectrum with respect to the total energy and the gray level of the image. The normalization process suppresses the influence of contrast, and we believe this is the cause of low image quality variance across age groups for MITRE.

We observed similar trends to Figures 3, 4 and 5 when we generated graphs for each device (details are in [26]).

[6] There are a few exceptions to this, such as 63 year old females and 57 year old males, where there is a clear increase in image quality (though the sample sizes in those cases are relatively low).

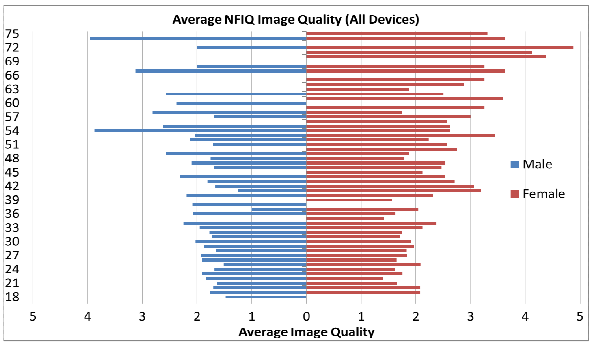


Figure 4. Christmas tree diagram showing the average NFIQ image quality (x-axis) pre age (y-axis) and gender (bar color) across all devices.

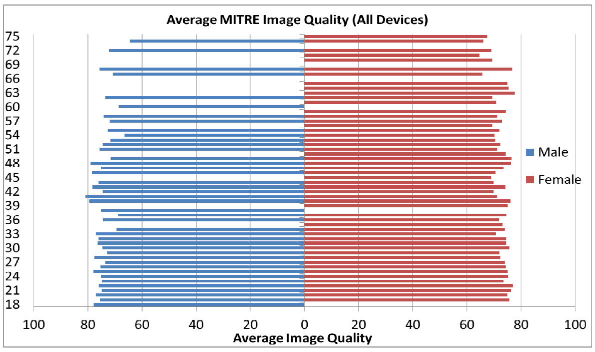


Figure 5. Christmas tree diagram showing the average MITRE image quality (x-axis) pre age (y-axis) and gender (bar color) across all devices.

2) *Image Quality Algorithm Comparison*

We also compared how the two fingerprint quality algorithms rate the quality of the same image. In an attempt to make the MITRE algorithm's quality scores more comparable with NFIQ, we grouped the MITRE scores into four broad groups[7]: 0-25 (1-lowest score), 26-50 (2), 51-75 (3) and 76-100 (4-highest score). Again, we generated graphs exhaustively for images from each device, subcategorized by age and gender (full details are in [26]).

Figure 6 is a sample from these exhaustive results. The plots show the distribution of fingerprint quality pairs for female participants' images captured with device D0, for each age group.

The x-axis shows the MITRE scores grouped into four categories as explained above, and the y-axis shows the NFIQ scores. The highest image quality pair (4 for MITRE, 1 for NFIQ) is in the top right corner. A value in the plot represents the frequency of images that were assigned a quality score from NFIQ (x-axis) and a score from MITRE (y-axis). A high frequency of images is represented by a dark red color and a low frequency by light yellow. Grey squares represent image quality pairs with zero images.

The plots illustrate that the MITRE algorithm does not discriminate image quality as effectively as NFIQ does, which is consistent with what we showed previously. In each of the plots, NFIQ assigns a quality score between, at least, 1 and 4. MITRE however, only assigns quality scores between 3 and 4. This is especially prominent in the 60+ age group, for both genders. In fact, we had a total of 34 images that had the worst NFIQ quality rating of 5, but were rated in the excellent group (76-100) by MITRE.

According to MITRE, 99.63% of the images in our dataset have a quality score in the range of 3 and 4 (50-100).

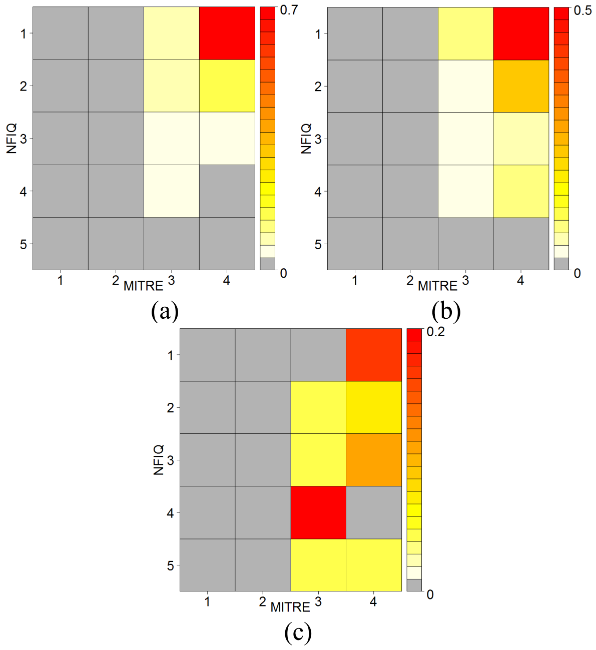
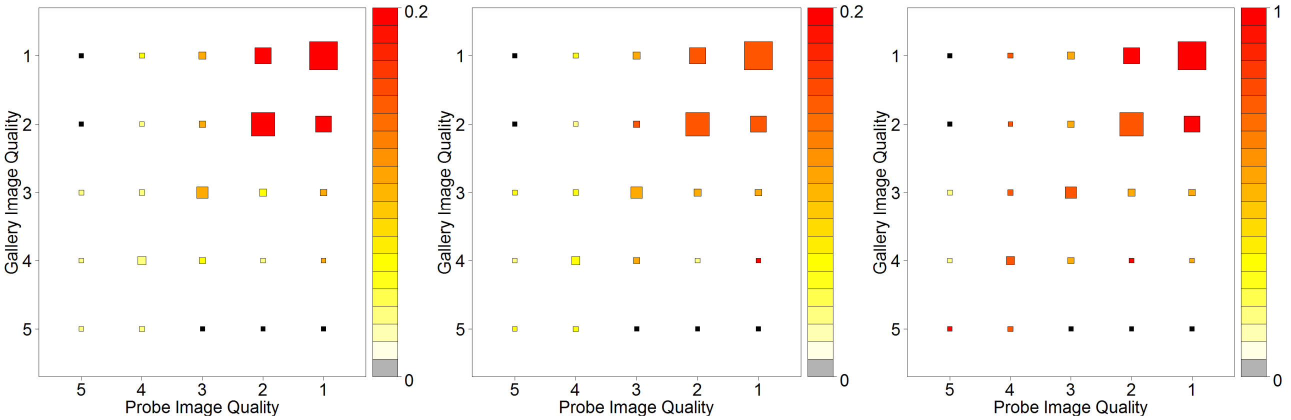


Figure 6. 3-Dimensional plots comparing the fingerprint quality scores assigned to female participants using capture device D0 by NFIQ (y-axis) and MITRE (x-axis). Plot (a) is with 18-29 year olds, (b) is with 30-59 year olds, and (c) is for 60+ year old participants.

Figure 7 contains three 4-dimensional plots showing how the NFIQ image quality of probe and gallery images, captured by device D3, affects the resulting genuine match scores from algorithms (a) M1, (b) M2 and (c) M3. The intensity of the color shows the normalized match score (light yellow=low, dark red=high and black=no data). The size of the square represents the proportion of data with a given image quality score pair, (probe image quality, gallery image quality). The size of the squares in each of the plots is the same because the image quality scores are independent of the matcher. The matching scores were normalized in the range 0 to the maximum matching score of each matcher (each matcher uses a different range).

Figure 7 (a) illustrates a clear correlation between high probe/gallery image quality scores resulting in a high match score from matcher M1. Figure 7 (b) shows that this trend also holds for matcher M2. It can also be seen from (a) and (b) that high match scores are still achievable when the probe image has a high image quality, such as with image quality pairs (1, 3) and (1, 4).

[7] These are based on the BioAPI quality score categories [27]



(a) (b) (c)

Figure 7. 4-Dimensional plots showing the average normalized genuine match score from matcher M1 (colour intensity) for comparsions with a certain probe image quality (x-axis) and gallery image quality (y-axis). The frequency of image quality pairs (x,y) is represented by the size of the square. Plot (a) uses match scores from matcher M1, (b) from matcher M2 and (c) from matcher M3.

Figure 7 (c) shows that matcher M3 is less sensitive to image quality: we can see that high match scores have also been achieved from low image quality pairs such as (5, 5), (4, 4) and (4, 5) as well as high image quality pairs. This is a desirable property of a matching algorithm.

We have generated plots exhaustively for each intra-device pair, matcher, match type (imposter/genuine), age and gender of participants (full details in [26]). For imposter graphs, in most cases we observe little variance in match scores across all probe/gallery quality pairs, except for device D0 with matcher M1 where we see a similar trend to Figure 7 (a). We have also noticed that swapping the probe and gallery images does not give the same match score i.e. the match scores are dependent on the order that images are passed to the matcher.

3) *Device Ranking by Image Quality*

Tables IV (a) and (b) summarize the rankings of the four devices by image quality per gender and age group (dark green=highest ranked; dark red=lowest ranked).

Table IV. The average (a) NFIQ and (b) MITRE image quality from each (age, gender, device) triplet.The final column shows the ordering of devices for each (age, gender) pair, based on a descending average image quality.

Participant		**Device**				**Device**
Age	**Gender**	**D0**	**D1**	**D2**	**D3**	**Sequence**
18-29	**Male**	1.530	1.878	1.905	1.702	D0, D3, D1, D2
	Female	1.396	1.935	2.104	1.735	D0, D3, D1, D2
30-59	**Male**	1.526	2.500	2.513	1.712	D0, D3, D1, D2
	Female	1.748	2.684	2.820	2.112	D0, D3, D1, D2
60+	**Male**	2.500	3.222	3.278	2.778	D0, D3, D1, D2
	Female	3.071	3.476	3.524	3.095	D0, D3, D1, D2

(a)

Participant		**Device**				**Device**
Age	**Gender**	**D0**	**D1**	**D2**	**D3**	**Sequence**
18-29	**Male**	83.530	74.449	73.199	71.577	D0, D1, D2, D3
	Female	82.087	73.761	71.617	71.483	D0, D1, D2, D3
30-59	**Male**	83.404	72.679	72.141	70.987	D0, D1, D2, D3
	Female	81.180	70.879	69.209	69.903	D0, D1, D3, D2
60+	**Male**	77.944	67.944	66.000	67.000	D0, D1, D3, D2
	Female	76.857	70.429	69.048	66.071	D0, D1, D2, D3

(b)

For both NFIQ (Table IV (a)) and MITRE (Table IV (b)) the device that captures the best quality images, on average, is D0 across genders and age groups. The two algorithms disagree on the second best device, with NFIQ listing D3 and MITRE D2. We explain in the next section whether the quality rankings agree with those from match scores.

B. *Fingerprint Match Analysis*

1) *Relationship of COTS Products by Match Scores*

As explained previously, we calculated match scores for fingerprint pairs using three different COTS products. Match scores were calculated under two setups: both gallery and probe images captured by the *same device* (*intra-device* match – DMG/DMI match scenarios, as described in Table II); and the probe and gallery images captured with *different devices* (*inter-device* matching – DDMG/DDMI match scenarios). For these setups, we calculated *genuine* match scores (i.e. when the gallery and probe images belong to the same participant – DMG/DDMG) and *impostor* match scores (when the gallery and probe images come from different participants – DMI/DDMI).

Figure 8 shows scatter plots of normalized match scores for the three COTS matcher pairs (a) (M1, M2), (b) (M1, M3) and (c) (M2, M3), where both probe and gallery images were captured with device D0 (intra-device). We will use Figure 8 (a) to explain what this and the subsequent plots are showing. The x-axis represents the normalized match score of COTS matcher M1 and the y-axis the normalized match score of matcher M2. The blue dots are the genuine match scores, and the red dots are the imposter match scores. For M1 and M2, we see that there is a positive correlation between the match scores, with M1 scores being generally higher[8].

Figures 8 (b) and (c) then show the comparison of match scores between COTS matcher pairs (M1, M3) and (M2, M3), respectively. Since most of the genuine scores (blue

[8] We had one very high match score from M1, as can be seen from the right hand side of Figure 8 (a). This causes the normalized scores of M1 to appear on the left half of the plots, but this does not affect the trend.

dots) fall on the top left hand corner of these graphs, this suggests that matcher M3 scores are generally higher than both M1 and M2. We see that the imposter scores are also generally higher for matcher M3, but there appears to be little overlap between genuine and imposter scores, suggesting that M3 outperforms both M1 and M2.

Figure 9 shows corresponding matcher pair scatter plots, where the gallery image was captured with device D0 and the probe image was captured with D3 (inter-device). Comparing Figures 8 and 9, we see that the trends are similar in general, but the genuine match scores, which indicate similarity, are lower in the inter-device plots (Figure 9).

Figure 10 (a) shows the bottom-left 10% of Figure 8 (a), which illustrates the trend of the imposter scores (red dots) and any overlapping areas with genuine scores. In general we see a trend towards higher imposter scores for matcher M2. Figures 8 (a) and 10 (a) seem to indicate that M1 outperforms M2: M1 scores are generally higher for genuine matches and lower for imposter matches, in comparison with M2. The actual performance of each COTS matcher will depend on where the matching threshold[9] is set.

Figure 10 (b) shows the 10% zoomed bottom-left corner of Figure 9 (a). Comparing Figures 10 (a) and (b) we observe:

- An increased number of genuine match score points visible in the zoomed inter-device plot, Figure 10 (b). This suggests that there are additional low genuine match scores in the inter-device scenario.
- An increased number of high imposter match scores in Figure 10 (b), compared with Figure 10 (a).
- A combination of low genuine match scores and high imposter match scores in the inter-device scenario would suggest that for these setups the performance would be negatively affected, due to the larger overlap.

We generated similar graphs for all other intra and inter device pairs, for all COTS product pairs; the trends we described for these figures are consistent in all cases (full details in [26]).

2) Device & COTS Matcher Ranking by Equal Error Rate

The performance of a biometric system in the identity verification mode is measured in terms of False Match Rates (FMR) and False Non-Match Rates (FNMR). The Equal Error Rate (EER) is the operating point at which the FMR and FNMR are equal, and it is commonly used in literature to compare the relative performance of different systems.

First we show the rankings of devices when both the probe and gallery images are captured with the same device (intra-device). Table V shows this ranking, sub-categorized by age, gender and matcher. The last column lists the device rankings, by ascending EER. In general, the trend is that device D0 has the lowest EER and is therefore ranked highest across the different subcategories.

The interesting point, however, is the existence of a variation in rankings related to the combination of human and system factors. For example, devices D2 and D3 are ranked highest in some situations. This is most common with female participants, which contributes to D3 being ranked highest with matcher M2 (second row in Table V).

More commonly with the older age groups, there are multiple devices with identical EER (where this is the case we have put these devices in curly brackets {}), though we should note that the number of participants in these groups is lower, as shown in section IV.A.

Table VI shows the EER for each COTS matcher for each device pair (both inter- and intra-device). We see that matcher M3 universally has the lowest EER. This is consistent with the results presented so far, where we saw that matcher M3 had higher genuine match scores compared with M1 and M2. Even though it had relatively high imposter scores, there seemed to be little overlap between them. This is now confirmed by the low EER in Table VI. In most cases, M1 is the second best matcher. This is again consistent with our observations in the previous section. We also see that in the majority of cases the intra-device configurations have the lowest EER.

Figure 11 shows the Equal Error Rate (EER) across all intra-device (a) and inter-device (b) pairs, categorized by participants' age group and gender, for each matching product. The main observations from Figure 11 are:

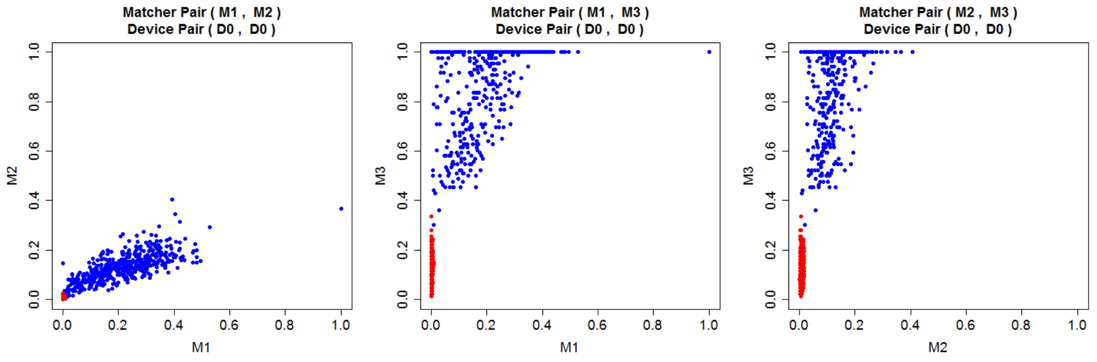
- Overall, the inter-device EERs are higher (worse) than those in the intra-device plots. There is also a steeper decline in performance for 18-29 and 30-59 age groups in the inter-device comparison scenarios.
- Both plots show a consistent increase (deterioration) in EER as the participant age group increases. There is a single exception to this, where the EER decreases between 30-59 and 60+ age groups for female participants with M3 matcher in intra-device scenario.
- Female participants generally have a higher (worse) EER than men.
- Matcher 3 is the least sensitive to the age of participants with one exception: where 30-59 year old females with matcher M1 have a lower EER than M3.

3) Optimal Verification Device with Fixed Enrollment

In the previous section we observed device rankings by EER. Here we look into the similarity match scores in more detail to analyze whether there are any special cases that would allow us to select an optimal device or matcher to maximize system performance. For each enrollment device, age group and gender, we ranked the highest match score with all verification devices. We then count the number of participants for whom a given verification device gives the highest match score.

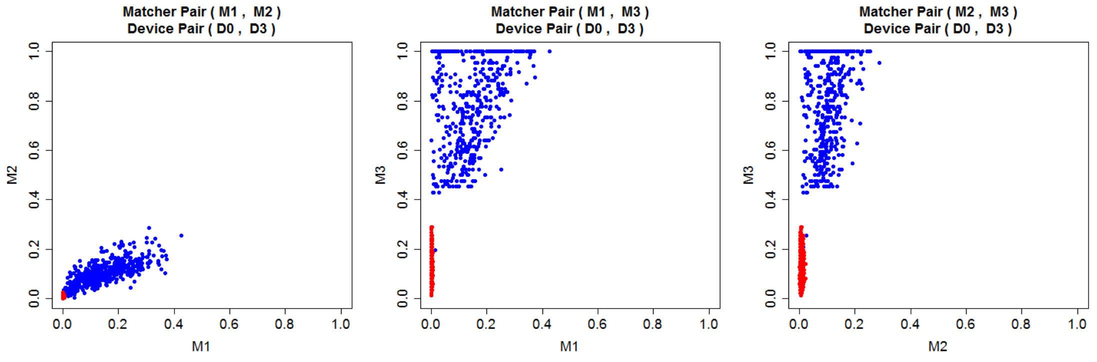
Table VII shows the proportion of highest ranked genuine match scores for each device pair and each matcher, categorized by age group and gender. To illustrate how these values should be read we use the first row, with COTS matcher M1, as an example. In this case the enrollment was carried out for 18-29 year old males with device D0. The value in the D0-D1 cell can be read as *"when we enrolled 18-29 males with device D0, for 13.69% of these participants, the highest match scores during verification were obtained with D1"*.

[9] The matching threshold is the value below which we would consider the matching outcome between the two fingerprints to be a reject.



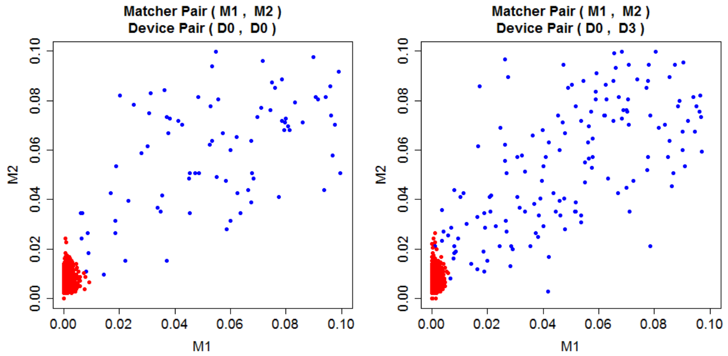
(a) (b) (c)

Figure 8. *Intra*-device scatter plots using probe/gallery images from device D0, showing the correlation of genuine (blue) and imposter (red) match scores between matchers (x-axis, y-axis) (a) (M1, M2), (b) (M1, M3), and (c) (M2, M3).



(a) (b) (c)

Figure 9. *Inter*-device scatter plots using (probe, gallery) images from devices (D0, D3), showing the correlation of genuine (blue) and imposter (red) match scores between matchers (x-axis, y-axis) (a) (M1, M2), (b) (M1, M3), and (c) (M2, M3).



(a) (b)

Figure 10. 10% zoomed scatter plots for matcher pair (M1, M2) for intra-device pair (a) (D0, D0) and inter-device pair (b) (D0, D3).

The percentages are grouped across verification devices for a given participant profile and enrollment device. The highest percentages for each grouping have been highlighted with a green cell background, second highest yellow, third highest orange, and last in red. Note that the percentages between the four verification devices for a given enrollment device do not always add up to 100%; one or more of the highest match scores may be identical across verification devices.

We notice a consistent trend in Table VII for COTS matchers M1 and M2, where the largest proportion of highest match scores are with intra-device pairs (the diagonal). We also notice that D2 as a verification device consistently gives the lowest proportion of highest match scores for inter-device pairs.

COTS matcher M3 differs from M1 and M2. We observe the following:

- D1 and D2 have identical proportion of highest match scores when used as the verification device.

- D0 has the highest proportion of match scores when the enrolment is carried out with device D0 or D2. (The only exception to this is for 60+ males, where the number of participants is small).
- For M3, as we illustrated with the scatter plots, we observe maximum match scores with multiple device pairings (a large number of blue dots in Figures 8 (b)-(c) and 9 (b)-(c) were in the upper most line). Hence, the proportions of highest match scores are much higher in comparison with matcher M1 and M2.

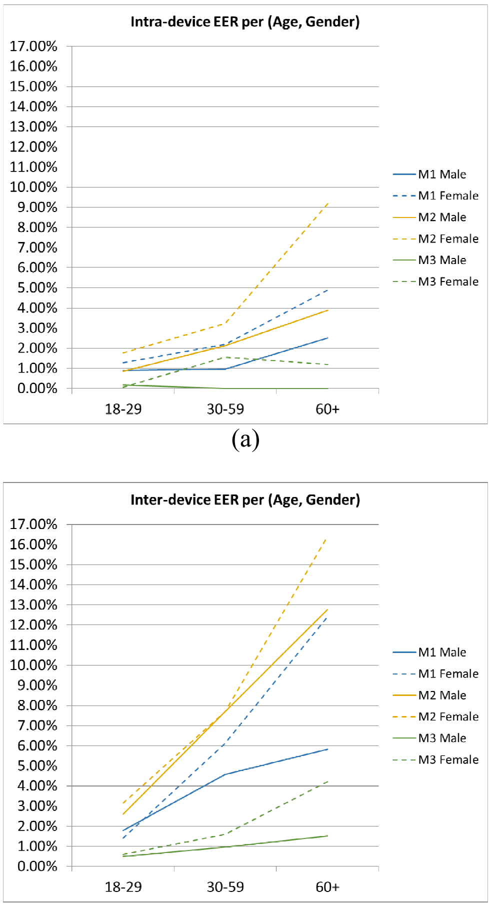
We have generated similar tables for imposter match scores, however, these have been omitted from the paper due to space constraints (full details are in [26]). In summary, we did not observe as clear a trend as we did for the genuine scores. The proportion of lowest imposter scores is more evenly spread between the verification devices. But, we did observe that for matchers M1 and M2, device D1 consistently had the highest proportion of lowest match scores.

Table V. Device performance sequences in an intra-device setup, for each (age, gender, matcher) subcategorization, using Equal Error Rate (EER) as a performance measure.

Age	**Gender**	**Matcher**	**Device Sequence (By EER, Ascending)**
ALL	ALL	1	D0,D3,D2,D1
ALL	ALL	2	D3,D0,D2,D1
ALL	ALL	3	D0,D3,D2,D1
18-29	MALE	1	D0,D1,D3,D2
18-29	MALE	2	D0,D1,D2,D3
18-29	MALE	3	{D0,D1,D2},D3
18-29	FEMALE	1	D3,D0,D2,D1
18-29	FEMALE	2	D0,D3,D2,D1
18-29	FEMALE	3	{D0,D1},{D2,D3}
30-59	MALE	1	{D0,D3},D2,D1
30-59	MALE	2	{D0,D3},D2,D1
30-59	MALE	3	{D0,D1,D2,D3}
30-59	FEMALE	1	D0,D3,D2,D1
30-59	FEMALE	2	D3,D0,D2,D1
30-59	FEMALE	3	{D0,D3},D1,D2
60+	MALE	1	{D0,D1,D2,D3}
60+	MALE	2	{D0,D1,D2},D3
60+	MALE	3	{D0,D1,D2,D3}
60+	FEMALE	1	D0,{D2,D3},D1
60+	FEMALE	2	D2,D3,D0,D1
60+	FEMALE	3	{D0,D2,D3},D1

Table VI. The Equal Error Rate calculated for each matcher for all device pairs.

Device		**Matcher (EER)**		
Probe	**Gallery**	**1**	**2**	**3**
D0	**D0**	0.399%	0.554%	0.003%
D0	D1	3.057%	5.106%	0.915%
D0	D2	2.607%	2.561%	0.704%
D0	D3	0.783%	1.158%	0.499%
D1	D0	2.840%	4.946%	0.939%
D1	**D1**	2.472%	4.409%	0.719%
D1	D2	2.952%	6.609%	0.735%
D1	D3	2.902%	5.316%	1.221%
D2	D0	2.044%	2.973%	0.538%
D2	D1	2.977%	6.824%	0.628%
D2	**D2**	1.896%	1.709%	0.608%
D2	D3	3.553%	3.560%	0.662%
D3	D0	0.806%	1.083%	0.383%
D3	D1	3.185%	5.420%	0.872%
D3	D2	3.812%	3.613%	0.859%
D3	**D3**	1.668%	1.100%	0.227%



(a)

(b)

Figure 11. Equal Error Rates (y-axis) from each matcher algorithm by age group (y-axis) and gender (dashed line=female, solid=male). Plot (a) is with all intra-device pairs, and (b) is with all inter-device pairs.

Table VII. Showing the percentage of DDMG scans (where participant$_{probe}$ == participant$_{gallery}$) with the highest match score for each verification device (columns) for a given participant profile (age, gender) and enrollment device (rows), from COTS products M1, M2 and M3.

Matcher / Verification Device (With highest % of genuine match scores)

Participant Profile / Enrolment Device			M1				M2				M3			
			D0	**D1**	**D2**	**D3**	**D0**	**D1**	**D2**	**D3**	**D0**	**D1**	**D2**	**D3**
18-29	**M**	**D0**	73.214	13.69	3.571	9.524	72.024	7.143	7.143	14.881	74.405	36.31	36.31	36.31
		D1	26.786	54.762	5.952	12.5	19.643	51.19	8.333	23.214	46.429	58.929	58.929	33.929
		D2	23.81	7.738	62.5	5.952	16.071	3.571	70.238	10.714	47.024	36.905	36.905	30.357
		D3	21.429	9.524	2.976	66.071	13.69	7.738	5.357	73.81	38.69	36.31	36.31	64.881
	F	**D0**	80.87	6.957	5.217	6.957	75.652	5.217	4.348	15.652	73.043	26.957	26.957	32.174
		D1	22.609	59.13	7.826	10.435	19.13	46.957	14.783	21.739	45.217	46.087	46.087	41.739
		D2	12.174	13.913	67.826	6.087	17.391	4.348	72.174	7.826	44.348	35.652	35.652	35.652
		D3	19.13	9.565	2.609	68.696	11.304	1.739	1.739	86.087	35.652	39.13	39.13	54.783
30-59	**M**	**D0**	73.077	8.974	5.128	12.821	70.513	6.41	5.128	17.949	71.795	29.487	29.487	29.487
		D1	19.231	55.128	7.692	21.795	28.205	55.128	3.846	19.231	41.026	51.282	51.282	41.026
		D2	16.667	8.974	66.667	7.692	14.103	8.974	62.821	14.103	52.564	46.154	46.154	34.615
		D3	12.821	5.128	3.846	78.205	15.385	6.41	5.128	75.641	33.333	33.333	33.333	62.821
	F	**D0**	78.641	7.767	1.942	11.65	69.903	5.825	2.913	25.243	59.223	48.544	48.544	31.068
		D1	17.476	59.223	11.65	12.621	22.33	58.252	11.65	16.505	52.427	55.34	55.34	49.515
		D2	16.505	12.621	61.165	9.709	19.417	5.825	67.961	9.709	48.544	48.544	48.544	43.689
		D3	26.214	2.913	0.971	69.903	13.592	4.854	2.913	80.583	33.981	41.748	41.748	54.369
60+	**M**	**D0**	55.556	11.111	0	33.333	88.889	0	11.111	0	33.333	44.444	44.444	55.556
		D1	0	55.556	33.333	11.111	22.222	55.556	0	22.222	44.444	88.889	88.889	33.333
		D2	22.222	11.111	55.556	11.111	22.222	0	66.667	11.111	33.333	44.444	44.444	22.222
		D3	11.111	0	22.222	66.667	22.222	11.111	11.111	66.667	44.444	44.444	44.444	33.333
	F	**D0**	66.667	9.524	4.762	19.048	66.667	4.762	9.524	28.571	61.905	23.81	23.81	47.619
		D1	33.333	47.619	9.524	9.524	19.048	47.619	14.286	23.81	42.857	57.143	57.143	33.333
		D2	14.286	4.762	76.19	4.762	23.81	4.762	71.429	4.762	52.381	38.095	38.095	42.857
		D3	23.81	4.762	0	71.429	14.286	0	0	85.714	47.619	23.81	23.81	61.905

V. DISCUSSION

In this section we build upon our exploratory analysis of fingerprint quality (IV.A) and similarity match scores (IV.B) to make evidence-based recommendations for optimizing fingerprint biometric system deployments.

A. Discussion of Fingerprint Quality Analysis

The main conclusions we can draw from the fingerprint quality analysis with our data set are:

- *Observation*: Image quality decreases with age for both genders. This is consistent across all devices for both image quality algorithms. This is to be expected as with age hand moisture levels decrease, the cumulative effect of finger injuries increases and our fingerprints deteriorate. Hence the quality of the fingerprint image also deteriorates.
  Recommendation: For older subjects, where the operating environments allow it, it is important to set a more stringent threshold for acceptable fingerprint image quality scores. This may cause delays in image capture (due to multiple acquisition attempts), but it would lead to improved authentication performance.
- *Observation:* Male participants exhibit higher fingerprint image quality. This may be explained by the fact that biometric devices have traditionally been most widely deployed in male-dominated operating environments, such as the military and law enforcement. Hence, we conjecture that the quality algorithms have also been optimized using male-oriented data sets. Additionally, as noted in [8] and [9], female fingerprints have higher ridge density. With more ridges in the same image area, ridge clarity might suffer and feature extraction is more difficult. This could have a negative impact on quality and subsequently on match scores.
  Recommendation: For operating environments where biometric devices are currently deployed, such as in border control, it is important that the system designers are aware that the fingerprint quality algorithms as well as the image capture software within the devices may have been optimized for a predominately young male population. This may necessitate setting the threshold on match scores conditional on the subjects' age and gender, provided one can trust these, to optimize system performance.
- *Observation:* The lowest fingerprint quality scores are obtained by older females (60+). This again may be explained by our conjecture on the training sets used to optimize fingerprint quality algorithms (young, predominantly male populations).
  Recommendation: Same as in previous bullet point.
- *Observation:* NFIQ is a better discriminator for fingerprint image quality than MITRE IQF.
  Recommendation: Given a choice between these two fingerprint quality algorithms, our analysis suggests that the system designer should consider using the NFIQ algorithm to better discriminate between good and bad quality fingerprint images. Accepting and using a good quality fingerprint image is important

since we have observed a positive correlation between image quality and match scores.

- *Observation:* D0 captures the best quality fingerprint images across all age groups and genders. There is an agreement on the best device by both NFIQ and MITRE, though they disagree on the second best device for image quality (NFIQ ranks D3 as second best; MITRE ranks D1).
  Recommendation: If the designer has a choice, then D0 would be the most optimal device amongst the tested ones, on average, for fingerprint capture in both enrollment and verification.

B. *Conclusions for COTS Matching Product Analysis*

The main conclusions we can draw from the COTS matching product analysis with our data set are:

- *Observation:* The match scores for intra-device pairings are generally better than for inter-device ones, especially for M1 and M2. However COTS matcher M3 seems to improve device interoperability.
  Recommendation: If the system architect has a choice, then the same device for enrollment and verification should be used. Otherwise, if the operational environment cannot avoid a diverse setup, product M3 should be used to mitigate against lower match scores that may result from using different devices.
- *Observation:* With COTS matchers M1 and M2, the largest proportion of highest match scores are obtained with intra-device configurations (Table VII). This would suggest that these matchers are highly sensitive to the choice of fingerprint capture devices (hence they would offer low interoperability). *Recommendation:* If there is a choice, as stated in the previous bullet point, M3 is, on average, the optimal COTS product to use when images have been captured with diverse devices.
- *Observation:* D0 tends to be the device with the highest match scores and lowest EERs on average. But there are some exceptions (usually for female participants) where devices D2 and D3 were ranked the highest.
  Recommendation: The most optimal device pair in the majority of cases, in terms of system performance, is where D0 is used for both enrollment and verification. But the optimal choice differs in some cases, especially for female participants cf. earlier discussion on the datasets used for algorithm optimization.
- *Observation:* M3 is the best COTS product in the majority of configurations we looked at (highest match scores, and lowest EERs), but there are some exceptions when we categorize by age and gender. It is also the product that is least affected by low quality fingerprints (as we noted in section IV.A), age, or inter-device matching (as we observe from Table VII), which would suggest that this is also the best COTS product, on average, to use to improve interoperability.

  Recommendation: If there is a choice, using matcher M3 will result, in the majority of cases, in optimal system performance. However, this matcher is also more expensive than the other two in our study.
- *Observation:* Matchers M1 and M2 assign the largest proportion of highest match scores to intra-device configurations (Table VII).
  Recommendation: In configurations where the same device is used for enrollment and verification, M1 and M2 can provide comparable and sometimes better performance than M3 (see technical report for full details [26]).

VI. CONCLUSIONS

In this paper we presented results of a large-scale empirical study in which the fingerprints of 494 participants were captured with 4 diverse fingerprint biometric devices. Match scores were generated with 3 different COTS matching products, two of which are commercial and one (M2) is available as open source. The fingerprint image quality was analyzed with 2 popular image quality algorithms. The results were then categorized per gender and age group. This allowed us to make a range of empirically-supported recommendations on the design and deployment principles of fingerprint biometric systems.

A summary of main conclusions is as follows.

- Better image quality fingerprints are obtained from:
  - Younger participants: this is mainly due to wear and tear, and physiological ageing processes in the fingerprint and is consistent with results from other studies [7], [8].
  - Male participants: we conjecture that this is because the fingerprint image quality algorithms, biometric capture devices and match algorithms in general tend to be optimized for male-oriented populations such as the military and law enforcement.
- NIST NFIQ fingerprint image quality algorithm is a better discriminator of quality compared with MITRE IFQ.
- Cross Match Guardian R2 fingerprint capture device (abbreviated as D0 in our paper) was on average the device that captured the best quality images, and from which the most optimal performance was obtained in terms of match scores and equal-error rates.
- On average, better (higher) match scores are obtained when the same device is used for image capture at both the enrollment and verification stages, indicating interoperability problems in fingerprint capture devices.
- BIO-key WEB-key Software Development Kit Matcher (abbreviated as M3 in our paper) was on average the matcher that gave the most optimal performance (in terms of equal error rates) and was the least affected by the source of the image capture devices, or the quality of the fingerprint image. So this would be the COTS matcher we would recommended to improve, on average, match scores when the fingerprint images have been captured with diverse devices. We should note that it is also a more

expensive of the three COTS products in our study.

It should be noted that even though we selected *best on average* fingerprint capture devices, fingerprint quality algorithms and match score algorithms, none of them could be rated as *best universally* in our data set. Hence a promising avenue for further work, which we are currently pursuing, is how we can take advantage of diversity at these layers to improve the overall system performance.

The results we presented apply to the studied devices, COTS matchers and fingerprint image quality algorithms but are likely to indicate broader concerns for the mass deployment of biometric authentication. However we refrain from making more detailed generalizations until we complete statistical significance tests. We are currently working on quantifying the significance related to the recommendations made. Another limitation of our study which prevents us from making more general conclusions is that the data was collected primarily at a University campus (at West Virginia University). Hence the convenience population sampling in the collection resulted in the majority of young college students.

Other avenues for further work include:

- Analyzing the effect of user habituation on image quality and match scores. In other words, if the users are instructed on how to provide a biometric sample with a specific device, does the image quality and match scores improve significantly? This might be especially important for intra-device deployments to improve interoperability and for female or elderly participants where we see the lowest image quality and match scores.
- Analyzing the different unimodal (i.e. fingerprints only) multi-device score fusion techniques to improve overall system performance of intra-device deployments.
- Analyzing the improvements that may be obtained from multimodal fusion (e.g. fingerprints as well as iris or face biometrics).
- Analyzing what soft-biometric traits can be inferred from fingerprint images. This might be especially useful in forensic applications.

ACKNOWLEDGMENT

This work was supported in part by the City University London Pump Priming fund via grant "Empirical Analysis of Diversity for Security". The work at West Virginia University was supported in part through the NIJ award # 2010-IJ-CX-K024 and NSF award IIP-1066197. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the sponsors.

REFERENCES

[1] D. Maltoni, D. Maio, A. Jain and S. Prabhakar. *Handbook of fingerprint recognition*. Springer, 2003.

[2] L. Lugini, E. Marasco, B. Cukic, I. Gashi, "Interoperability in Fingerprint Recognition: a Large Scale Study", *The 43rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks* (DSN) - Workshops, June 2013, Budapest, Hungary.

[3] E. Marasco, L. Lugini, B. Cukic, T. Bourlai, "Minimizing the Impact of Low Interoperability between Optical Fingerprint Sensors", Biometrics: Theory, Applications and Systems (BTAS), pp. 1-8, Washington DC, Sept. 29 - Oct. 2, 2013.

[4] N. Poh, J. Kittler, and T. Bourlai. "Quality-based score normalization with device qualitative information for multimodal biometric fusion", *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 40.3 (2010): 539-554.

[5] A. Ross and A. Jain. "Biometric sensor interoperability: A case study in fingerprints." *Biometric Authentication*. Springer Berlin Heidelberg, 2004. pp. 134-145.

[6] A. Ross and R. Nadgir,"A calibration model for fingerprint sensor interoperability", *Proceedings of SPIE*, 6202, 2006.

[7] A. Jain, S. Dass, and K. Nandakumar, "Can soft biometric traits assist user recognition?", *Defense and Security*, *International Society for Optics and Photonics*, pp. 561–572, 2004.

[8] S. Modi, S. Elliott, J. Whetsone, and H. Kim., "Impact of age groups on fingerprint recognition performance", *IEEE Workshop on Automatic Identification Advanced Technologies*, pages 19–23, 2007.

[9] Y. Nigeria. Towards age & gender determination, ridge thickness to valley thickness ratio (RTVTR) & ridge count on gender detection. Analysis, Design and Implementation of Human Fingerprint Patterns System, 1(2), 2012.

[10] M. Acree. Is there a gender difference in fingerprint ridge density? *Forensic science international*, 102(1):35–44, 1999.

[11] Badawi, A., Mahfouz, M., Tadross, R., & Jantz, R. (2006, June). Fingerprint-Based Gender Classification. In IPCV (pp. 41-46).

[12] Nithin, M. D., Manjunatha, B., Preethi, D. S., & Balaraj, B. M., "Gender differentiation by finger ridge count among South Indian population", *Journal of forensic and legal medicine*, 18(2), 79-81.

[13] P. Gnanasivam, and S. Muttan. "Estimation of Age Through Fingerprints Using Wavelet Transform and Singular Value Decomposition." *International Journal of Biometrics and Bioinformatics* (IJBB) 6.2 (2012): 58-67.

[14] E. Marasco, " Secure Multibiometric Systems", *PhD Dissertation*., Università degli Studi di Napoli Federico II, 2010.

[15] A. Jain, and A. Ross. "Multibiometric systems." *Communications of the ACM,* 47.1 (2004): 34-40.

[16] K. Nandakumar, " Multibiometric systems: Fusion strategies and template security", ProQuest, 2008.

[17] A. Ross, and A. Jain. "Information fusion in biometrics." *Pattern recognition letters* 24.13 (2003): 2115-2125.

[18] A. Jain, P. Flynn, A. Ross, *Handbook of biometrics*. Springer, 2008.

[19] L. Kuncheva, "Combining Pattern Classifiers Method and Algorithms", Wiley, 2004.

[20] Y. Ma, B. Cukic, and H. Singh, "A classification approach to multi-biometric score fusion", In T. Kanade, A.K. Jain, and N.K. Ratha, editors, AVBPA, volume 3546 of *Lecture Notes in Computer Science*, pages 484–493. Springer, 2005.

[21] K. Nandakumar, Y. Chen, S. Dass, "Likelihood ratio-based biometric score fusion", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30.2 (2008): 342-347.

[22] E. Marasco, A. Ross, and C. Sansone. "Predicting identification errors in a multibiometric system based on ranks and scores." *The Fourth IEEE International Conference on Biometrics: Theory Applications and Systems* (BTAS), IEEE, 2010.

[23] C. Watson, M. Garris, E. Tabassi, C. Wilson, R. McCabe, and S. Janet. Users guide to NIST fingerprint image software 2 (nfis2). National Institute of Standards and Technology, 2004.

[24] E. Marasco and C. Sansone, "Improving the accuracy of a score fusion approach based on likelihood ratio in multimodal biometric systems", *Lecture Notes in Computer Science*, 5716:509–518, 2009.

[25] G. Marcialis and F.Roli. "Serial fusion of fingerprint and face matchers", *Lecture Notes in Computer Science*,4472:151–160,June 2007

[26] S. Mason, I. Gashi, "Deployment Strategies for Diverse Fingerprint Biometric Systems: Technical Report", http://www.csr.city.ac.uk/people/ilir.gashi/DSN2014/

[27] N. Nill, "IQF (Image Quality of Fingerprint) Software Application". *http://www.mitre.org/sites/default/files/pdf/07_0580.pdf*