Benchmarking Fingerprint Minutiae Extractors

Tarang Chugh[1], Sunpreet S. Arora[2,*], Anil K. Jain[1], and Nicholas G. Paulter Jr.[3]

[1]Department of Computer Science and Engineering, Michigan State University, East Lansing, Michigan 48824

[2] Emerging Technology, Risk and Authentication Products Group, Visa Inc., Foster City, CA, 94404

[3] National Institute of Standards and Technology, 100 Bureau Dr., Gaithersburg, MD, 20899

E-mail: [1]{chughtar,jain}@cse.msu.edu, [2]sunarora@visa.com, [3]paulter@nist.gov

***Abstract*—The performance of a fingerprint recognition system hinges on the errors introduced in each of its modules: image acquisition, preprocessing, feature extraction, and matching. One of the most critical and fundamental steps in fingerprint recognition is robust and accurate minutiae extraction. Hence we conduct a repeatable and controlled evaluation of one open-source and three commercial-off-the-shelf (COTS) minutiae extractors in terms of their performance in minutiae detection and localization. We also evaluate their robustness against controlled levels of image degradations introduced in the fingerprint images. Experiments were conducted on (i) a total of $3,458$ fingerprint images from five public-domain databases, and (ii) $40,000$ synthetically generated fingerprint images. The contributions of this study include: (i) a benchmark for minutiae extractors and minutiae interoperability, and (ii) robustness of minutiae extractors against image degradations.**

***Index Terms*—fingerprint recognition, minutiae extraction, robustness to noise, interoperability**

I. Introduction

A fingerprint recognition system typically comprises of four major modules: image acquisition, preprocessing, feature extraction, and matching (See Fig. 1). The errors introduced in each of these four modules, from image acquisition to matching cumulatively impact the overall system recognition performance. For instance, the low fidelity[4] of a fingerprint signal acquired by a sensor can introduce errors in preprocessing, induce poor feature extraction, and ultimately deteriorate the matching performance. Therefore, it is important to perform a comprehensive evaluation of each module independently to improve the overall performance of the fingerprint recognition system.

Fingerprint sensor certification standards (*e.g. PIV-071006* [1] and *Appendix F* [2]) mandate independent evaluation of fingerprint sensors. Hence vendors are required to demonstrate that their sensors can acquire a high-fidelity image with low-noise characteristics. Existing studies have evaluated the performance of sensors in terms of their resilience to external environmental factors (temperature and humidity), intrinsic subject-dependent factors (skin humidity and pressure) [3], operational quality [4], their interoperability [5], and finger liveness detection [6]. Arora et. al [7] have designed and fabricated 3D fingerprint targets and whole hand targets for repeatable evaluation and calibration of fingerprint sensors.

On the contrary, studies pertaining to fingerprint preprocessing, feature extraction, and matching, evaluate these modules in entirety as a black-box with the goal to improve the overall matching performance. National Institute of Standards and Technology (NIST) conducts fingerprint vendor technology evaluations (FpVTE) to benchmark the capabilities of fingerprint recognition systems in terms of identification accuracy and computational requirements [8], [9]. The 2014 FpVTE [9] reports that the best performing system achieved a FNIR of 1.9% for single index finger, and 0.09% using all ten-fingers, at a FPIR of 0.1%. Fingerprint verification competitions[5] (FVC 2000-2006) also evaluate systems from an end-to-end perspective. Although these third-party evaluations are useful, they do not evaluate individual modules. For instance, in the case of a false match or a non-match, it is uncertain whether the error is caused due to poor image quality, minutiae extraction errors, or inability of the matcher to handle distortion. An independent evaluation of the individual modules will enable us to understand the error sources and design an interoperable system.

It is generally known that minutiae extraction is critical to fingerprint recognition accuracy. Minutiae-based representation is the most widely used approach, essentially due to its (i) interpretability, (ii) high matching performance, (iii) storage efficiency, (iv) applicability to match fingerprints/latents in forensic casework, and (v) evidential value (*i.e.* expert testimony based on mated minutiae is admissible in the courts of law) [10]. The FVC-onGoing [11], in addition to benchmarking performance at the system level, also provides benchmarks for (i) fingerprint orientation extraction, and (ii) matching standard minutiae-based templates [ISO/IEC 19794-2 (2005)]. However, accuracy and robustness evaluation of minutiae extracted using different minutiae extractors are needed in order to benchmark their performance and minutiae interoperability.

Minutiae interoperability tests (*e.g.* MINEX III [12]) evaluate the compliance between minutiae-based template generators and matchers from different vendors. Kayaoglu et al. [13] compared the matching performance based on automatically extracted minutiae and manually labelled minutiae. However,

This research was supported by grant no. 60NANB11D155 from the NIST Measurement Science program.

*At the time this research was conducted, Sunpreet was affiliated with the Dept. of Computer Science and Engineering, Michigan State University.

[4]Fidelity refers to the degree of exactness with which friction ridge patterns on a finger are reproduced by the sensor

[5]https://biolab.csr.unibo.it/FVCOnGoing/UI/Form/Home.aspx

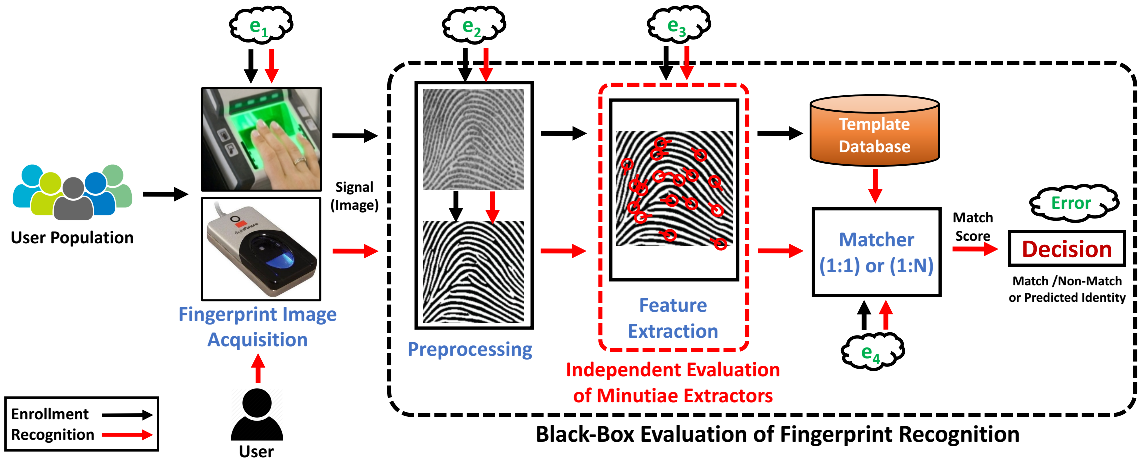

Fig. 1: Framework of a typical fingerprint recognition system. While existing studies evaluate the recognition system from an end-to-end perspective, we provide a benchmark for minutiae extraction module. Errors introduced at different steps of the system, *i.e.* fingerprint acquisition (e_1), preprocessing (e_2), minutiae extraction (e_3), and matching (e_4), cumulatively impact the overall performance.

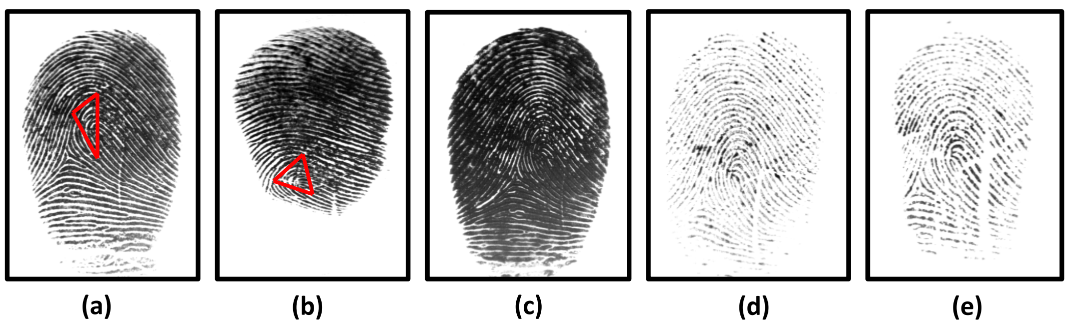

Fig. 2: Challenges in automated fingerprint processing. Five different impressions of the same finger (from FVC2004 DB1A). These illustrate (a) reference fingerprint, (b) large non-linear distortion (compare the triangle in (b) to triangle in the reference fingerprint (a)), (c) smudged areas due to wet fingerprint, (d) and (e) broken ridge structure due to dry and noisy fingerprints.

these tests did not evaluate the underlying factors limiting the minutiae interoperability, *i.e.* variations in the minutiae detection and localization ability. Moreover, the images input to minutiae extractors may contain distortion and motion blur due to variance in pressure applied on the sensor platen, and may have poor contrast due to dry/wet fingers (See Fig. 2). To address these challenges, this study conducts:

- A repeatable and controlled evaluation of minutiae extraction in terms of their detection and localization performance, for one open-source and three commercial minutiae extractors.
- A rigorous assessment of robustness of minutiae extractors in the presence of controlled levels of noise and motion blur to understand their limitations.

II. Evaluation Protocol

A. Databases

The fingerprint images used in this evaluation study are grouped into two sets.

- Dataset-A contains $3,458$ real fingerprint images compiled from five public domain databases: FVC 2002 (DB1A and DB3A), FVC 2004 (DB1A and DB3A) and NIST SD27 rolled prints database[6]. Each FVC database contains 800 fingerprint images (100 unique subjects, 8 acquisitions/subject), with ground truth minutiae marked by human subjects [13]. NIST SD27 [16] contains 258 rolled prints with ground truth minutiae marked by at least two certified forensic examiners.
- Dataset-B contains $40,000$ synthetic fingerprints (including $5,000$ unique masterprints, and $35,000$ fingerprints degraded with controlled levels of noise and motion blur) generated using Novetta's biosynthetic software [17]. It contains four levels of noise (including anatomical deformations, dryness, ridge noise) and three levels of motion blur.

[6]NIST SD27 is no longer publicly available.

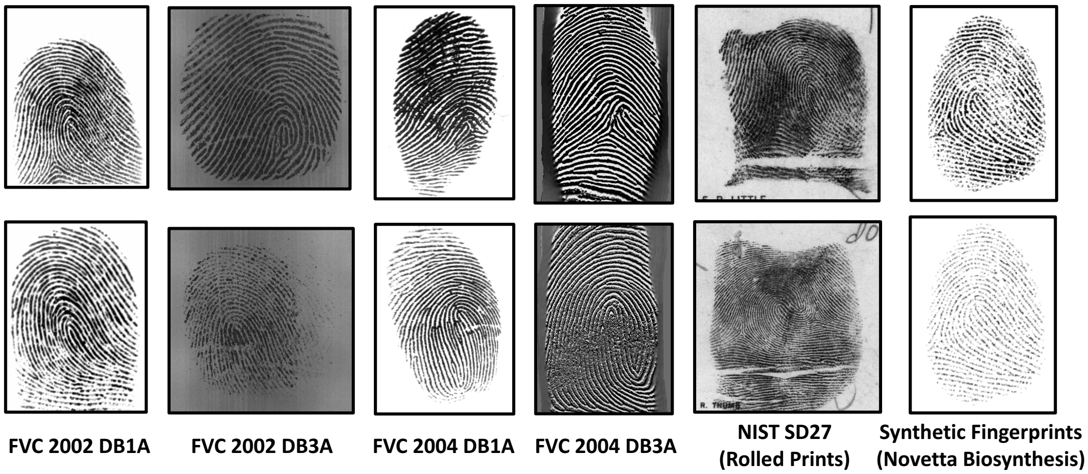

Fig. 3: Examples of fingerprint images from the six databases used in this evaluation study.

Database	**(# Fingerprints, # Subjects)**	**Ground Truth**	**Image Capture**	**Image Size ($h \times w$)**	**Avg. NFIQ2 value (s.d.)**
		Dataset-A			
FVC2002 DB1A [14]	(800, 100)	Manually Marked Minutiae	Optical sensor	374×388	64 (15)
FVC2002 DB3A [14]	(800, 100)		Capacitive sensor	300×300	26 (13)
FVC2004 DB1A [15]	(800, 100)		Optical sensor	480×640	59 (17)
FVC2004 DB3A [15]	(800, 100)		Thermal sweep sensor	480×300	47 (16)
NIST SD27 (rolled prints) [16]	(258, 258)		Digitized ink and paper	768×800	42 (10)
		Dataset-B			
Synthetic masterprints [17]	(5,000, 5,000)	N/A	Synthetically generated	480×512	71 (6)
Noisy prints [17]	(20,000, 5,000)	Minutiae extracted from master prints	Synthetically generated	480×512	40 (23)
Motion blurred prints	(15,000, 5,000)		Synthetically generated	480×512	44 (26)

TABLE I: A summary of fingerprint databases used in this evaluation study.

Figure 3 presents example fingerprint images from each of these databases. The two sets of fingerprint databases used in this study are summarized in Table I. The average NIST Fingerprint Image Quality 2.0 (NFIQ 2.0) [18], which lies in the range $[0, 100]$ where 0 indicates the worst quality, and 100 refers to the best quality, is also presented for each database.

B. Evaluating Minutiae Detection and Localization

An ideal fingerprint minutiae extractor is expected to exhibit high precision in minutiae detection and localization, and minimize spurious and missing minutiae. We evaluate the performance of one open-source minutiae extractor *mindtct* [19], and three minutiae extractors (COTS - A, B, and C) by comparing the extracted minutiae with the ground truth obtained from human subjects for Dataset-A. The performance of a fingerprint minutiae extractor depends heavily on the quality of input fingerprint images. Considering the large variations in the NFIQ 2.0 values, we segregate the fingerprint images from Dataset-A into five quality bins $[0, 20]$, $[21, 40]$, $[41, 60]$, $[61, 80]$, and $[81, 100]$ based on the NFIQ 2.0 values. Figure 4 presents examples of fingerprint images corresponding to each of the 5 quality bins. For a fair evaluation, performance comparison between minutiae extractors is done only for fingerprint images within each quality bin. We do not utilize the synthetic fingerprint images (Dataset-B) for this evaluation, as the synthesis process itself introduces some spurious minutiae.

1) Minutiae Detection: Given a fingerprint image, let $F_d = \{f_d^1, f_d^2, ..., f_d^N\}$ be the set of N minutiae detected by a minutiae extractor, and $F_g = \{f_g^1, f_g^2, ..., f_g^M\}$ be the set of M ground truth minutiae marked by human subjects. A detected minutia f_d, and a ground truth minutia f_g are said to be *paired*, if f_d lies within a distance threshold δ around f_g. As the average ridge width for a 500 ppi fingerprint image is known to be approximately 9 pixels [20], we fix the threshold to 10 pixels. If there is more than one detected minutia within the threshold, the one closest to the ground truth minutia is paired with it. In case of a tie, the pairing decision is made in favor of the minutia with smaller orientation difference. If a minutia has to be inserted in the set F_d, in order to pair it with a minutia in the set F_g, it is considered as a *missing* minutia. Similarly, if a minutiae in the detected set F_d, cannot be paired with any minutia in ground truth set F_g, it is deemed to be a *spurious* minutia. We utilize the Goodness Index (GI) metric of Ratha et al. [21] to evaluate the minutiae detection performance.

$$GI = \frac{\sum_{i=1}^{L} Q_i [P_i - D_i - I_i]}{\sum_{i=1}^{L} Q_i M_i} \tag{1}$$

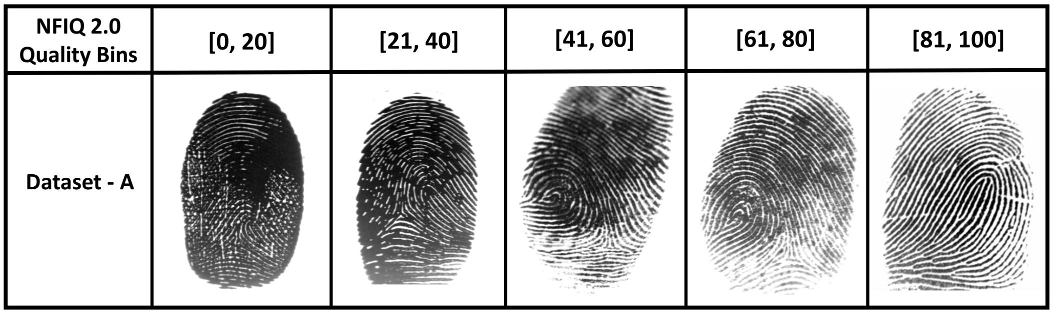

Fig. 4: Examples of fingerprint images from Dataset-A corresponding to the 5 quality bins based on NFIQ 2.0 values, where $[0, 20]$ represents the worst quality bin and $[81, 100]$ indicates the best quality bin.

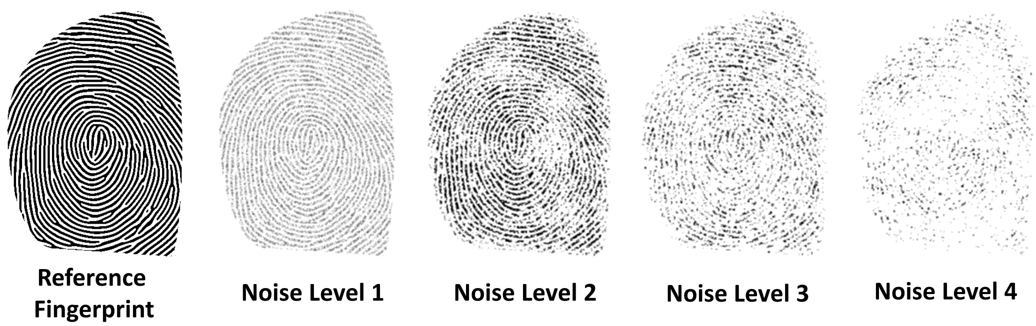

Fig. 5: Four different levels of noise added to the master fingerprint (reference fingerprint).

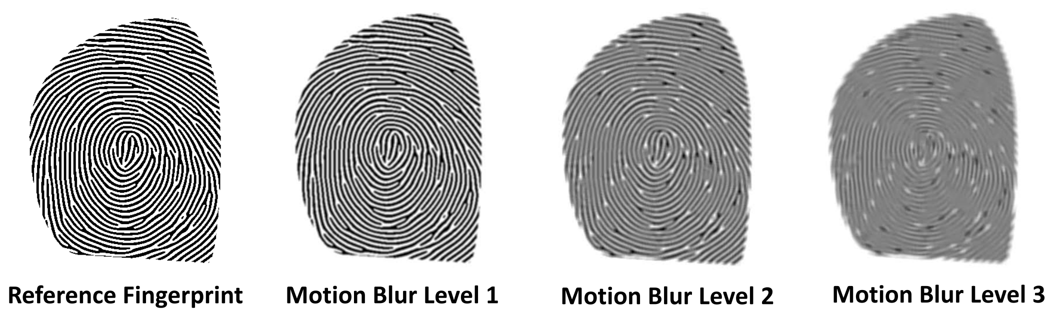

Fig. 6: Three different levels of motion blur added to the master fingerprint (reference fingerprint).

where L = no. of 16×16 non-overlapping patches in the input image, Q_i = quality of the i^{th} patch ($good = 4,\ medium = 2,\ poor = 1$), P_i = no. of paired minutiae in the i^{th} patch, D_i = no. of spurious minutiae in the i^{th} patch, $D_i \leq 2 \cdot M_i$, I_i = no. of missing minutiae in the i^{th} patch, and M_i = no. of ground truth minutiae in the i^{th} patch, $M_i > 0$. In order to restrict the negative impact of outlier patches, the number of spurious minutiae (D_i) in a patch is restricted to a maximum value of $2 \cdot M_i$.

The quality index proposed by Chen et al. [22] is utilized. We do not consider patches with zero minutiae (near image boundary). The maximum value of GI is $+1$, which is obtained when $D_i = I_i = 0$ and $P_i = M_i$, *i.e.* all detected minutiae are paired and no. of detected and ground truth minutiae is the same. The minimum value of GI is -3, which is obtained when $P_i = 0$, $D_i = 2 \times M_i$, and $I_i = M_i$, *i.e.* no detected minutiae could be paired and the no. of spurious minutiae takes its maximum possible value of $2 \cdot M_i$. Larger the value of Goodness Index, better the performance of a minutiae extractor. In addition to Goodness Index (GI), we also report the average percentages of paired (P_i/M_i), spurious (D_i/M_i), and missing (I_i/M_i) minutiae.

2) Minutiae Localization: For a given minutiae extractor, let $\hat{f}_d = \{\hat{f}_d^1, \hat{f}_d^2, ..., \hat{f}_d^P\}$, $\hat{f}_d \subseteq F_d$, be a set of P detected minutiae points, paired with a subset of known ground truth minutiae points $\hat{f}_g \subseteq F_g$. The positional error (e_p) for the paired minutiae set ($\hat{f}_g, \hat{f}_d$) is computed using the Root Mean

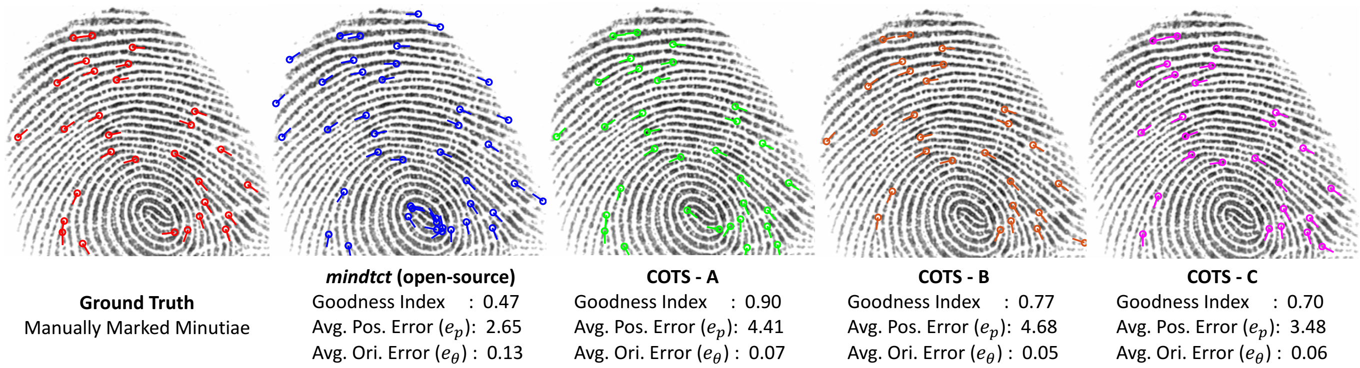

Fig. 7: Example fingerprint from FVC2002 DB1A dataset with overlaid manually marked minutiae and minutiae extracted by four minutiae extractors (*mindtct*, and COTS A, B, and C). Goodness Index (GI) is unit less, while Avg. Positional Error (e_p) and Avg. Orientation Error (e_θ) are measured in pixels and radians, respectively.

Square Deviation (RMSD) [23] given by:

$$e_p(\hat{f}_g, \hat{f}_d) = \sqrt{\frac{\sum_{i=1}^{P}[(x_g^i - x_d^i)^2 + (y_g^i - y_d^i)^2]}{P}} \tag{2}$$

where, (x_d^i, y_d^i) and (x_g^i, y_g^i) represent the locations of the detected minutia and the ground truth minutia, respectively. Similarly, the orientation error (e_θ) between the set of paired minutiae $(\hat{f}_g, \hat{f}_d)$ is computed using:

$$e_\theta(\hat{f}_g, \hat{f}_d) = \sqrt{\frac{\sum_{i=1}^{P}\phi(\theta_g^i, \theta_d^i)^2}{P}} \tag{3}$$

where

$$\phi(\theta_1, \theta_2) = \begin{cases} \theta_1 - \theta_2 & \text{if} \quad -\pi \le \theta_1 - \theta_2 < \pi \\ 2\pi + \theta_1 - \theta_2 & \text{if} \quad \theta_1 - \theta_2 < -\pi \\ -2\pi + \theta_1 - \theta_2 & \text{if} \quad \theta_1 - \theta_2 \ge \pi \end{cases}$$

C. Evaluating Robustness of Minutiae Extractors

The primary reason of errors in minutiae detection is the presence of artifacts due to variations in finger placement on the sensor platen, noise, finger moisture, fingerprint alterations, etc. A common evaluation technique, known as *stress testing*, is used to test a system beyond normal operating conditions, often to a breaking point. We evaluate the robustness of one open-source minutiae extractor *mindtct* [19], and three commercial minutiae extractors in the presence of controlled levels of noise, finger dryness, and motion blur, to understand the stable operational conditions. We utilize the synthetic fingerprint images from Dataset-B for this evaluation.

1) Robustness against Noise: Fingerprint images acquired by the fingerprint readers may possess noise due to physical factors such as anatomical deformations in the friction ridge skin (scars, holes, scratches, etc.), finger moisture, and/or environmental contamination. These noise sources induce significant variation in minutiae extraction, even within multiple acquisitions of the same finger. To quantify the impact of noise on minutiae extractors, synthetic prints with controlled levels of noise are generated from synthetic master fingerprints. The noise model in Novetta's biosynthetic software [17] is utilized to add (i) anatomical deformations (scars, holes, and pressure variations), (ii) ridge noise (Perlin noise), and (iii) finger dryness. Fig. 5 presents different levels of noise added to a master fingerprint (used as the reference).

2) Robustness against Motion Blur: Movements of the hand during fingerprint acquisition may lead to introduction of motion blur in the acquired image. We simulate three levels of motion blur in the synthetic master fingerprints by applying motion lens filter function in both horizontal and vertical direction [24]. The MATLAB functions $fspecial('motion', k)$ and $fspecial('motion', k, 90)$, with three different values of $k \in \{5,\ 7,\ \text{and}\ 9\}$ corresponding to increasing degrees of motion blur, are applied. Fig. 6 presents a synthetic master print and corresponding three different levels of motion blur.

III. Experimental Results

Goodness index, average positional error (e_p), and average orientation error (e_θ) are computed by comparing the output from one open-source minutiae extractor, *mindtct*, and three COTS minutiae extractors with the manually marked minutiae for Dataset-A, and minutiae extracted on the master print (without any image degradations) for Dataset-B.

A. Minutiae Detection and Localization

Fig. 7 presents an example fingerprint from FVC2002 DB1A dataset with overlaid manually marked minutiae and the extracted minutiae from one open-source minutiae extractor, *mindtct*, and three COTS minutiae extractors. The values for the three performance metrics, Goodness Index, Positional Error, and Orientation Error are also reported for each minutiae extractor output. Tab. II presents a summary of the performance comparison between the four minutiae extractors in terms of minutiae detection and localization accuracies for Dataset-A. In comparison to other minutiae extractors, COTS-B consistently achieves a higher value of Goodness Index across all quality levels. Performance of COTS-A is observed to be highly dependent on fingerprint quality, as it achieves the lowest Goodness Index for low quality images (NFIQ 2.0 =

NFIQ 2.0 Quality Bins	Minutiae Extractor	[0, 20]	[21, 40]	[41, 60]	[61, 80]	[81, 100]
# Fingerprints		419	803	$1,051$	$1,053$	132
Goodness Index Avg. (s.d.)	*mindtct*	-0.64 (0.77)	-0.45 (0.70)	-0.33 (0.59)	0.11 (0.38)	0.36 (0.25)
	COTS-A	-0.74 (0.69)	-0.14 (0.71)	0.00 (0.67)	0.47 (0.26)	0.60 (0.16)
	COTS-B	-0.03 (0.63)	0.22 (0.44)	0.33 (0.30)	0.48 (0.22)	0.57 (0.17)
	COTS-C	-0.04 (0.70)	0.12 (0.51)	0.21 (0.35)	0.40 (0.21)	0.48 (0.19)
Positional Error (e_p) **(in pixels)** Avg. (s.d.)	*mindtct*	3.95 (0.80)	3.78 (0.69)	3.60 (0.73)	3.22 (0.56)	3.10 (0.46)
	COTS-A	4.87 (0.66)	4.64 (0.61)	4.37 (0.64)	4.27 (0.60)	4.22 (0.59)
	COTS-B	4.53 (0.83)	4.24 (0.72)	4.02 (0.73)	4.00 (0.61)	3.89 (0.54)
	COTS-C	4.10 (0.86)	4.21 (0.82)	4.23 (0.78)	3.83 (0.70)	3.59 (0.57)
Avg. Orientation Error (e_θ) **(in rad.)** Avg. (s.d.)	*mindtct*	0.27 (0.23)	0.20 (0.12)	0.18 (0.09)	0.15 (0.06)	0.14 (0.04)
	COTS-A	0.16 (0.12)	0.13 (0.07)	0.12 (0.06)	0.11 (0.04)	0.10 (0.03)
	COTS-B	0.13 (0.13)	0.10 (0.06)	0.10 (0.05)	0.10 (0.04)	0.09 (0.03)
	COTS-C	0.14 (0.12)	0.11 (0.07)	0.10 (0.05)	0.10 (0.04)	0.09 (0.02)

TABLE II: Performance comparison of four minutiae extractors (*mindtct*, and COTS A, B, and C) in terms of minutiae detection and localization accuracies. This evaluation utilizes fingerprint images (Dataset-A) from five public domain datasets, available with manually marked ground truth minutiae. Minutiae detection is measured in terms of Goodness Index (GI), a unit less measure in the range [-3, 1]. A large value of GI suggests high number of detected minutiae are paired with ground truth minutiae and low number of spurious or/and missing minutiae.

NFIQ 2.0 Quality Bins	Minutiae Extractor	[0, 20]	[21, 40]	[41, 60]	[61, 80]	[81, 100]
# Fingerprints		419	803	$1,051$	$1,053$	132
Paired Minutiae / Ground Truth (P_i / M_i) Avg. (s.d.)	*mindtct*	0.77 (0.12)	0.81 (0.11)	0.82 (0.09)	0.84 (0.08)	0.86 (0.07)
	COTS-A	0.77 (0.14)	0.79 (0.16)	0.78 (0.17)	0.85 (0.07)	0.86 (0.06)
	COTS-B	0.71 (0.15)	0.76 (0.12)	0.79 (0.10)	0.82 (0.08)	0.84 (0.07)
	COTS-C	0.74 (0.14)	0.74 (0.11)	0.75 (0.09)	0.77 (0.08)	0.78 (0.09)
Spurious Minutiae / Ground Truth (D_i / M_i) Avg. (s.d.)	*mindtct*	1.19 (0.63)	1.06 (0.60)	0.97 (0.53)	0.57 (0.34)	0.36 (0.21)
	COTS-A	1.29 (0.60)	0.72 (0.52)	0.56 (0.44)	0.22 (0.20)	0.12 (0.09)
	COTS-B	0.44 (0.45)	0.30 (0.31)	0.25 (0.21)	0.15 (0.13)	0.10 (0.08)
	COTS-C	0.52 (0.55)	0.36 (0.39)	0.30 (0.28)	0.13 (0.12)	0.09 (0.08)
Missing Minutiae / Ground Truth (I_i / M_i) Avg. (s.d.)	*mindtct*	0.23 (0.12)	0.19 (0.11)	0.18 (0.09)	0.16 (0.08)	0.14 (0.07)
	COTS-A	0.23 (0.14)	0.21 (0.16)	0.22 (0.17)	0.15 (0.07)	0.14 (0.06)
	COTS-B	0.29 (0.15)	0.24 (0.12)	0.21 (0.10)	0.18 (0.08)	0.16 (0.07)
	COTS-C	0.26 (0.14)	0.26 (0.11)	0.25 (0.09)	0.23 (0.08)	0.22 (0.09)

TABLE III: Performance comparison of the four minutiae extractors (*mindtct*, and COTS A, B, and C) in terms of average percentages of paired (P_i/M_i), spurious (D_i/M_i), and missing (I_i/M_i) minutiae for fingerprint images of different quality (Dataset-A).

[0, 20]), and highest Goodness Index for high quality images (NFIQ 2.0 = [81,100]). The open-source minutiae extractor, *mindtct*, achieves low Goodness Index compared to COTS minutiae extractors across all quality values, however, it also achieves lowest positional errors suggesting high positional accuracy for the paired minutiae. In general, a NFIQ 2.0 quality value lower than 20 leads to a negative Goodness Index and higher localization errors with larger variances. It can be observed that as the quality level increases, the Goodness Index values also increase, indicating higher number of paired minutiae and lower number of spurious and/or missing minutiae. Tab. III presents the performance comparison of the four minutiae extractors in terms of average percentages of paired (P_i/M_i), spurious (D_i/M_i), and missing (I_i/M_i) minutiae. It can be observed that the open-source minutiae extractor produces a much higher percentage of spurious minutiae, but a much lower percentage of missing minutiae, compared to other COTS minutiae extractors.

B. Robustness against Image Degradations

Tab. IV summarizes the performance comparison between the four minutiae extractors on robustness against different levels of image noise for Dataset-B. It can be observed that as the noise level increases, the Goodness Index decreases, and the avg. positional error and the avg. orientation error increases. In comparison to other minutiae extractors, COTS-A achieves a much higher Goodness Index, and low positional and orientation errors even in the presence of higher levels of image noise. All the minutiae extractors exhibit similar avg. positional errors, but a much higher variance is observed in the case of COTS-C. Tab. V presents the performance comparison between the four minutiae extractors in terms of average percentages of paired (P_i/M_i), spurious (D_i/M_i), and missing (I_i/M_i) minutiae for images with different levels of

Noise Levels	Minutiae Extractor	Level 1	Level 2	Level 3	Level 4
Goodness Index Avg. (s.d.)	*mindtct*	0.36 (0.27)	0.09 (0.32)	-0.43 (0.33)	-0.80 (0.25)
	COTS-A	0.80 (0.12)	0.72 (0.14)	0.52 (0.21)	0.15 (0.37)
	COTS-B	0.53 (0.19)	0.43 (0.21)	0.19 (0.23)	-0.15 (0.30)
	COTS-C	0.72 (0.19)	0.53 (0.28)	-0.08 (0.44)	-0.60 (0.35)
Positional Error (e_p) **(in pixels)** Avg. (s.d.)	*mindtct*	2.27 (0.59)	2.87 (0.72)	3.86 (0.72)	4.55 (1.05)
	COTS-A	2.07 (0.55)	2.54 (0.61)	3.43 (0.67)	4.17 (0.73)
	COTS-B	2.11 (0.63)	2.75 (0.74)	3.80 (0.69)	4.54 (0.72)
	COTS-C	2.24 (0.64)	2.85 (0.79)	3.84 (0.91)	4.82 (2.02)
Avg. Orientation Error (e_θ) **(in rad.)** Avg. (s.d.)	*mindtct*	0.06 (0.04)	0.09 (0.07)	0.19 (0.14)	0.36 (0.30)
	COTS-A	0.03 (0.02)	0.04 (0.03)	0.06 (0.05)	0.13 (0.12)
	COTS-B	0.04 (0.02)	0.05 (0.03)	0.07 (0.06)	0.13 (0.12)
	COTS-C	0.03 (0.02)	0.04 (0.03)	0.07 (0.07)	0.14 (0.25)

TABLE IV: Robustness evaluation of four minutiae extractors (*mindtct*, and COTS A, B, and C) against different levels of noise (Dataset-B).

Noise Levels	Minutiae Extractor	Level 1	Level 2	Level 3	Level 4
Paired Minutiae / Ground Truth $(P_i$ / $M_i)$ Avg. (s.d.)	*mindtct*	0.75 (0.12)	0.63 (0.11)	0.42 (0.09)	0.24 (0.08)
	COTS-A	0.92 (0.14)	0.88 (0.16)	0.81 (0.17)	0.70 (0.07)
	COTS-B	0.78 (0.15)	0.74 (0.12)	0.64 (0.10)	0.51 (0.08)
	COTS-C	0.89 (0.14)	0.80 (0.11)	0.52 (0.09)	0.24 (0.08)
Spurious Minutiae / Ground Truth $(D_i$ / $M_i)$ Avg. (s.d.)	*mindtct*	0.14 (0.06)	0.18 (0.09)	0.27 (0.13)	0.28 (0.12)
	COTS-A	0.04 (0.04)	0.05 (0.04)	0.10 (0.08)	0.24 (0.18)
	COTS-B	0.03 (0.03)	0.04 (0.04)	0.09 (0.07)	0.17 (0.10)
	COTS-C	0.05 (0.05)	0.08 (0.06)	0.11 (0.08)	0.08 (0.08)
Missing Minutiae / Ground Truth $(I_i$ / $M_i)$ Avg. (s.d.)	*mindtct*	0.25 (0.12)	0.37 (0.14)	0.58 (0.13)	0.76 (0.12)
	COTS-A	0.08 (0.06)	0.12 (0.07)	0.19 (0.08)	0.30 (0.12)
	COTS-B	0.22 (0.09)	0.26 (0.09)	0.36 (0.10)	0.49 (0.12)
	COTS-C	0.11 (0.09)	0.20 (0.13)	0.48 (0.22)	0.76 (0.19)

TABLE V: Performance comparison of the four minutiae extractors (*mindtct*, and COTS A, B, and C) in terms of average percentages of paired (P_i/M_i), spurious (D_i/M_i), and missing (I_i/M_i) minutiae for fingerprint images with different levels of noise (Dataset-B).

noise. It can be observed that COTS-A achieved a very high percentage of paired minutiae and much lower percentage of missing minutiae, resulting in a high Goodness Index. In terms of spurious minutiae, *mindtct* is observed to consistently perform poorly across all noise levels compared to the COTS minutiae extractors, producing much higher percentage of spurious minutiae.

The performance comparison of the four minutiae extractors in terms of minutiae detection and localization accuracies for images degraded with different levels of motion blur is presented in Tab. VI. It is observed that COTS-A achieves high Goodness Index value compared to other minutiae extractors with low avg. positional and orientation errors. In general, higher level of motion blur results in large negative values of Goodness Index for all minutiae extractors. Tab. VII presents the performance comparison in terms of average percentages of paired (P_i/M_i), spurious (D_i/M_i), and missing (I_i/M_i) minutiae for images with different levels of motion blur. With increase in the motion blur levels, a much higher percentage of missed minutiae is observed compared to paired and spurious minutiae.

IV. Conclusions

Minutiae extraction is one of the most critical component of an automatic fingerprint identification systems. We have presented a controlled and repeatable evaluation of one open-source and three COTS minutiae extractors. Our experiments involve five public domain databases with manually marked minutiae to determine minutiae detection and localization accuracies. A large synthetically generated database with controlled levels of image degradations allowed us to quantify the affects of noise and motion blur, on minutiae extraction performance. The open-source minutiae extractor (*mindtct*) is observed to produce lowest positional errors in public domain databases. However, it also generates a higher percentage of spurious minutiae compared to COTS minutiae extractors, deteriorating its overall performance. COTS-A exhibits significantly high robustness against different levels of image noise and motion blur.

References

[1] N. B. Nill, "Test Procedures for Verifying Image Quality Requirements for Personal Identity Verification (PIV) Single Finger Capture Devices," *MITRE, MTR 060170*, 2006.

Motion Blur Levels	Minutiae Extractor	Level 1	Level 2	Level 3
Goodness Index Avg. (s.d.)	*mindtct*	0.76 (0.12)	0.40 (0.16)	−0.68 (0.24)
	COTS-A	0.90 (0.13)	0.48 (0.16)	−0.50 (0.25)
	COTS-B	0.81 (0.15)	0.51 (0.15)	−0.56 (0.17)
	COTS-C	0.88 (0.10)	0.46 (0.13)	−0.70 (0.26)
Positional Error (e_p) **(in pixels)** Avg. (s.d.)	*mindtct*	3.05 (0.19)	3.69 (0.37)	4.14 (0.35)
	COTS-A	3.13 (0.20)	3.73 (0.38)	4.09 (0.31)
	COTS-B	3.08 (0.22)	3.84 (0.47)	4.10 (0.40)
	COTS-C	3.11 (0.19)	3.88 (0.31)	4.27 (0.58)
Avg. Orientation Error (e_θ) **(in rad.)** Avg. (s.d.)	*mindtct*	0.02 (0.01)	0.06 (0.02)	0.10 (0.02)
	COTS-A	0.01 (0.00)	0.06 (0.02)	0.09 (0.02)
	COTS-B	0.01 (0.01)	0.04 (0.01)	0.10 (0.03)
	COTS-C	0.01 (0.00)	0.06 (0.01)	0.08 (0.02)

TABLE VI: Robustness evaluation of four minutiae extractors (*mindtct*, and COTS A, B, and C) against different degrees of motion blur (Dataset-B).

Motion Blur Levels	Minutiae Extractor	Level 1	Level 2	Level 3
Paired Minutiae / Ground Truth $(P_i$ / $M_i)$ Avg. (s.d.)	*mindtct*	0.90 (0.09)	0.73 (0.14)	0.26 (0.18)
	COTS-A	0.96 (0.08)	0.76 (0.15)	0.34 (0.16)
	COTS-B	0.93 (0.09)	0.78 (0.14)	0.30 (0.16)
	COTS-C	0.95 (0.07)	0.75 (0.15)	0.25 (0.17)
Spurious Minutiae / Ground Truth $(D_i$ / $M_i)$ Avg. (s.d.)	*mindtct*	0.04 (0.03)	0.06 (0.04)	0.20 (0.13)
	COTS-A	0.02 (0.01)	0.04 (0.03)	0.18 (0.11)
	COTS-B	0.05 (0.03)	0.05 (0.04)	0.16 (0.13)
	COTS-C	0.02 (0.02)	0.04 (0.03)	0.20 (0.12)
Missing Minutiae / Ground Truth $(I_i$ / $M_i)$ Avg. (s.d.)	*mindtct*	0.10 (0.04)	0.27 (0.08)	0.74 (0.26)
	COTS-A	0.04 (0.02)	0.24 (0.06)	0.66 (0.19)
	COTS-B	0.07 (0.02)	0.22 (0.05)	0.70 (0.24)
	COTS-C	0.05 (0.02)	0.25 (0.06)	0.75 (0.20)

TABLE VII: Performance comparison of the four minutiae extractors (*mindtct*, and COTS A, B, and C) in terms of average percentages of paired (P_i/M_i), spurious (D_i/M_i), and missing (I_i/M_i) minutiae for fingerprint images with different levels of motion blur (Dataset-B).

[2] N. B. Nill, "Test Procedures for Verifying IAFIS Image Quality Requirements for Fingerprint Scanners and Printers," *MITRE Tech. Rep. MTR 05B0000016*, 2005.

[3] H. Kang *et al.*, "A study on performance evaluation of fingerprint sensors," in *ICAVBPA*, pp. 574–583, Springer, 2003.

[4] R. Cappelli, M. Ferrara, and D. Maltoni, "On the operational quality of fingerprint scanners," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 2, pp. 192–202, 2008.

[5] F. Alonso-Fernandez *et al.*, "Dealing with sensor interoperability in multi-biometrics: The UPM experience at the Biosecure Multimodal Evaluation 2007," in *SPIE Defense and Security Symposium*, pp. 69440J–69440J, 2008.

[6] L. Ghiani *et al.*, "LivDet 2013 Fingerprint Liveness Detection Competition 2013," in *ICB*, pp. 1–6, 2013.

[7] S. S. Arora, K. Cao, A. K. Jain, and N. G. Paulter, "Design and Fabrication of 3D Fingerprint Targets," *IEEE Trans. on Info. Forens. and Secur.*, vol. 11, no. 10, pp. 2284–2297, 2016.

[8] C. Wilson *et al.*, "Fingerprint Vendor Technology Evaluation 2003: Summary of Results and Analysis Report," *NIST Research Report*, vol. 7123, 2004.

[9] C. Watson *et al.*, "Fingerprint Vendor Technology Evaluation," *NIST Interagency Report 8034*, 2014.

[10] A. K. Jain, J. Feng, and K. Nandakumar, "Fingerprint matching," *Computer*, vol. 43, no. 2, 2010.

[11] B. Dorizzi *et al.*, "Fingerprint and On-Line Signature Verification Competitions," in *ICB*, pp. 725–732, 2009.

[12] P. Flanagan, "Minutiae Interoperability Exchange III (MINEX III)," *NIST*, 2015.

[13] M. Kayaoglu, B. Topcu, and U. Uludag, "Standard fingerprint databases: Manual minutiae labeling and matcher performance analyses," *arXiv preprint arXiv:1305.1443*, 2013.

[14] D. Maio *et al.*, "FVC2002: Second Fingerprint Verification Competition," in *ICPR*, vol. 3, pp. 811–814, 2002.

[15] D. Maio *et al.*, "FVC2004: Third Fingerprint Verification Competition," in *ICBA*, pp. 1–7, 2004.

[16] "NIST SD27 Latent Database." https://www.nist.gov/srd/nistsd27.cfm.

[17] "Novetta Biosynthetic Software," 2014. https://www.novetta.com/wp-content/uploads/2014/11/NOV_Biosynthetics_Overview-2.pdf.

[18] "National Institute of Standards and Technology, Development of NFIQ 2.0," 2016. https://www.nist.gov/services-resources/software/development-nfiq-20.

[19] "National Institute of Standards and Technology (NIST) Biometric Image Software (NBIS)," 2016. https://www.nist.gov/services-resources/software/nist-biometric-image-software-nbis.

[20] D. Maltoni *et al.*, *Handbook of Fingerprint Recognition*. Springer Science & Business Media, 2009.

[21] N. K. Ratha, S. Chen, and A. K. Jain, "Adaptive Flow Orientation-based Feature Extraction in Fingerprint Images," *Pattern Recognition*, vol. 28, no. 11, pp. 1657–1672, 1995.

[22] Y. Chen, S. C. Dass, and A. K. Jain, "Fingerprint quality indices for predicting authentication performance," in *ICAVBPA*, pp. 160–170, Springer, 2005.

[23] F. Turroni *et al.*, "Improving fingerprint orientation extraction," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 1002–1013, 2011.

[24] X. Liu, M. Pedersen, C. Charrier, P. Bours, and C. Busch, "The influence of fingerprint image degradations on the performance of biometric system and quality assessment," in *Biometrics Special Interest Group (BIOSIG), 2016 International Conference of the*, pp. 1–6, IEEE, 2016.