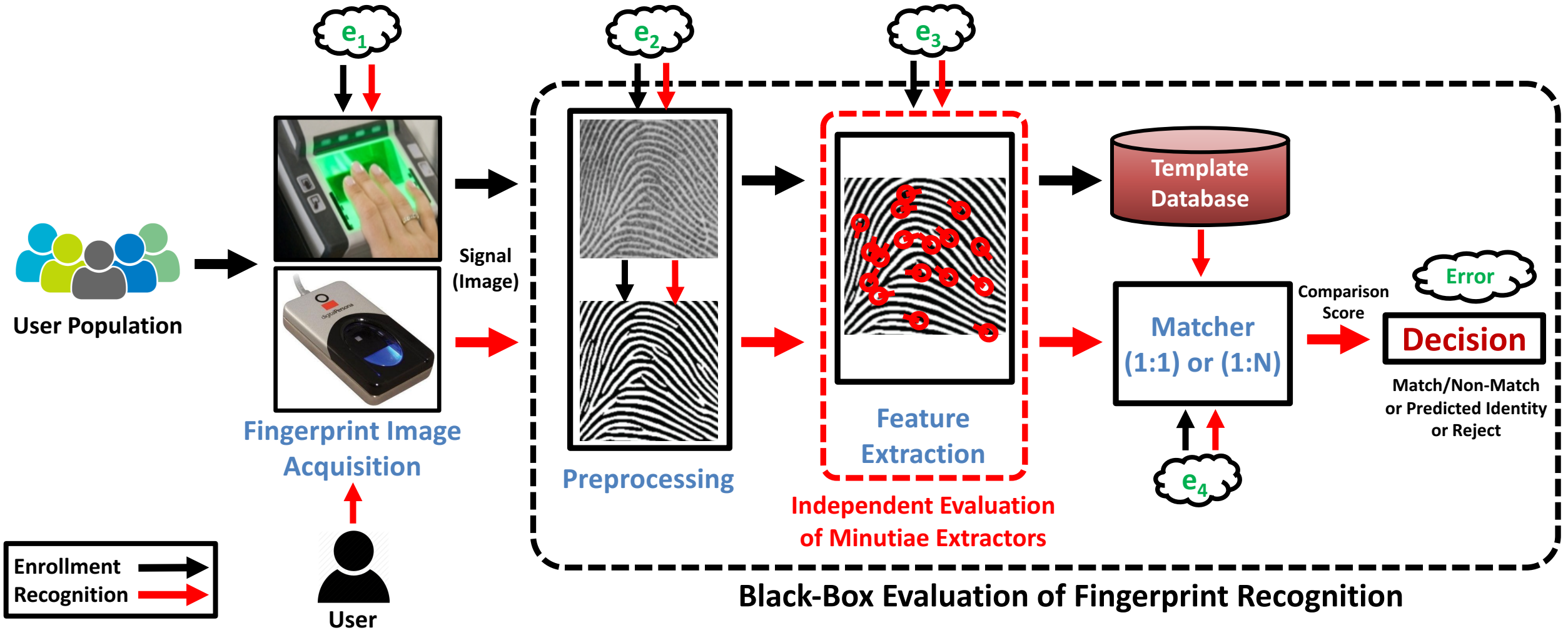
Benchmarking Fingerprint Minutiae Extractors

Tarang Chugh [†], Sunpreet S. Arora[*], Anil K. Jain[†] and Nicholas G. Paulter Jr.[‡]

[†] Michigan State University
[*]Visa Inc.
[‡] National Institute of Standards and Technology, U.S.A.

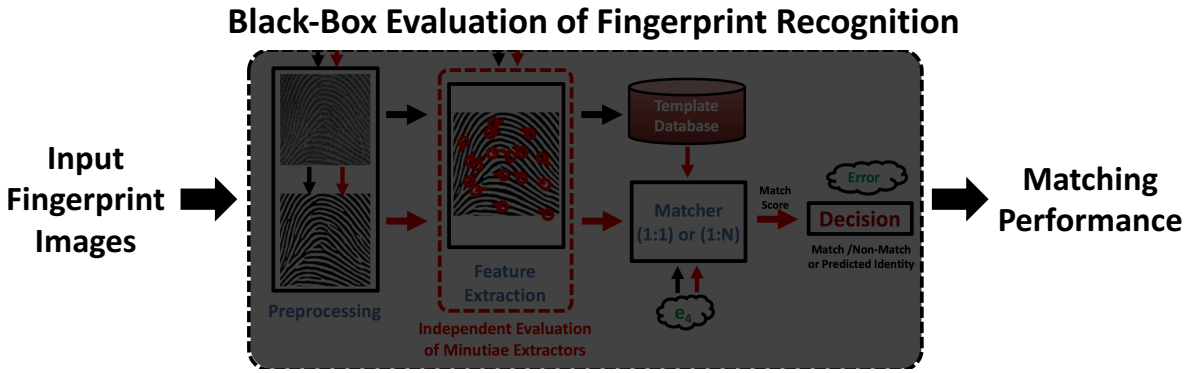
September 22, 2017

Fingerprint Recognition System



Black-box Evaluations of Fingerprint Recognition

- **Goal:** Evaluate overall matching performance (end-to-end perspective)
- NIST Fingerprint Vendor Technology Evaluations (FpVTE) [1]
  - Benchmark overall identification accuracy (FpVTE2014: FNIR = 0.09% @ FPIR = 0.1% for 10 fingers)
  - No evaluation of individual modules
- Cause of false match/non-match is uncertain;
  - Poor image quality?
  - Feature extraction errors?
  - Matcher unable to handle distortion?

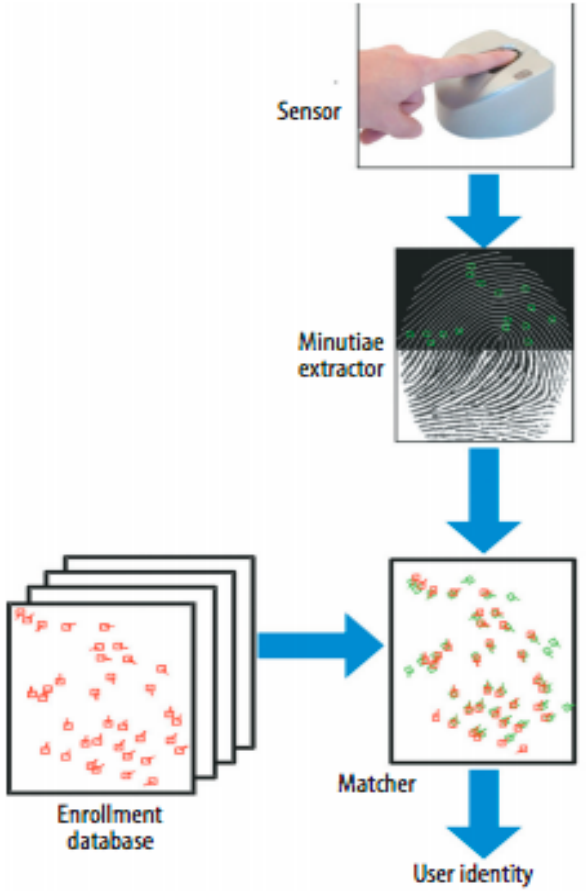


Component level evaluations will enable us to **identify and address error sources** and **design interoperable systems**.

[1] Watson, C. et al.: Fingerprint Vendor Technology Evaluation. NIST Interagency Report 8034, 2014.

Why Minutiae?

- Widely used Minutiae-based representation [1]
  1. Interpretability,
  2. Standardized,
  3. Compact Representation,
  4. High matching performance,
  5. Applicability to match fingerprints/latents in forensic casework, and
  6. Evidential value (expert testimony is admissible in courts of law)



Using minutiae-based representation to match fingerprints. Image from [10]

[1] Jain, A. K.; Feng, J.; Nandakumar, K.: Fingerprint matching. Computer, 43(2), 2010.

Existing Minutiae-based Evaluations

1. FVC-onGoing [1] benchmarks:
   - fingerprint orientation extraction, and
   - matching standard minutiae-based templates [ISO/IEC 19794-2 (2005)]
2. Minutiae interoperability tests (e.g. MINEX III [2])
   - Evaluates compliance between minutiae-based template generators and matchers
3. Kayaoglu et al. [3] compared matching performance based on automatically extracted and manually labelled minutiae

Underlying factors limiting the minutiae interoperability are not evaluated, i.e. variations in the **minutiae detection** and **localization**

[1] Dorizzi, B. et al.: Fingerprint and On-Line Signature Verification Competitions. In: ICB. pp. 725–732, 2009.
[2] Flanagan, P.: Minutiae Interoperability Exchange III (MINEX III). NIST, 2015.
[3] Kayaoglu, M.; Topcu, B.; Uludag, U.: Standard fingerprint databases: Manual minutiae labeling and matcher performance analyses. arXiv preprint arXiv:1305.1443, 2013.

Challenges in Minutiae Extraction

Five different impressions of the same finger (from FVC2004 DB1A)

Reference Fingerprint

Non-linear distortion

Smudged areas (wet fingers)

Broken ridge structure due to dryness and noise

Need for benchmarking **accuracy** and **robustness** of minutiae extraction

Contributions

- Repeatable and controlled evaluation of minutiae extraction in terms of their detection and localization performance, for one open-source (*mindtct [1]*) and three commercial minutiae extractors (*COTS A, B, and C*).

- Rigorous assessment of robustness of minutiae extractors in the presence of controlled levels of noise and motion blur to understand their limitations.

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

Databases

Database	(#Fingerprints, #Subjects)	Ground Truth	Image Capture	Image Size (h x w)	Avg. NFIQ2 value (s.d.)
Dataset-A					
FVC2002 DB1A [1]	(800, 100)	Manually Marked Minutiae	Optical sensor	374 x 388	64 (15)
FVC2002 DB3A [1]	(800, 100)		Capacitive sensor	300 x 300	26 (13)
FVC2004 DB1A [1]	(800, 100)		Optical sensor	480 x 640	59 (17)
FVC2004 DB3A [1]	(800, 100)		Thermal sweep sensor	480 x 300	47 (16)
NIST SD27 [2] (rolled prints)	(258, 258)		Digitized ink and paper	768 x 800	42 (10)
Dataset-B					
Synthetic masterprints [3]	(5,000, 5,000)	N/A	Synthetically Generated	480 x 512	71 (06)
Noisy prints [3]	(20,000, 5,000)	Minutiae extracted from master prints			40 (23)
Motion blurred prints [3]	(15,000, 5,000)				44 (26)

[1] Kayaoglu, M.; Topcu, B.; Uludag, U.: Standard fingerprint databases: Manual minutiae labeling and matcher performance analyses. arXiv preprint arXiv:1305.1443, 2013.
[2] NIST SD27 Latent Database, https://www.nist.gov/srd/nistsd27.cfm.
[3] Novetta Biosynthetic Software, https://www.novetta.com/wpcontent/uploads/2014/11/NOV Biosynthetics Overview-2.pdf.

Databases

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

FVC 2002 DB1A | **FVC 2002 DB3A** | **FVC 2004 DB1A** | **FVC 2004 DB3A** | **NIST SD27 (Rolled Prints)** | **Synthetic Fingerprints (Novetta Biosynthesis)**

Evaluation Protocol

1. Evaluate Minutiae Detection and Localization
   - Utilize Dataset-A with manually marked minutiae as ground truth
   - Fingerprints segregated into 5 bins based on NFIQ 2.0 [1] quality values

NFIQ 2.0 Quality Bins	[0, 20]	[21, 40]	[41, 60]	[61, 80]	[81, 100]
Dataset - A					
# fingerprints	419	803	1,051	1,053	132

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

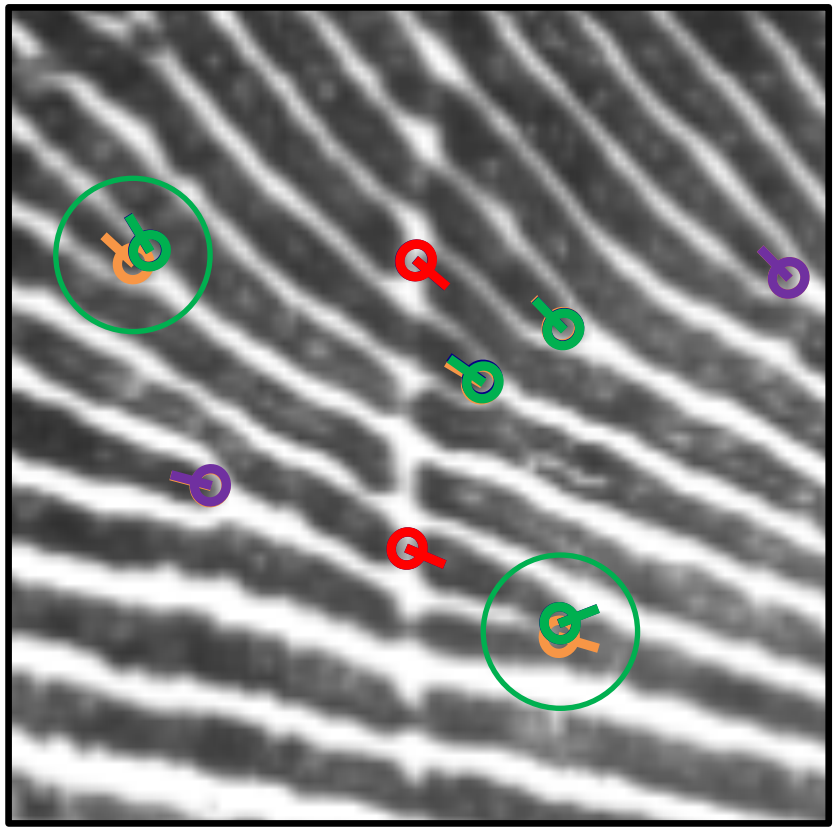
Evaluation Protocol

2. Evaluate robustness against Noise and Motion Blur
   - Utilize Dataset-B with controlled levels of noise and motion blur added to high quality synthetic master fingerprints
   - Minutiae extracted from the master fingerprint are considered as ground truth

Reference Fingerprint | **Noise Level 1** | **Noise Level 2** | **Noise Level 3** | **Noise Level 4**

Terminology

- **Ground Truth Minutiae:** $F_g = \{f_g^1, f_g^2 \ldots, f_g^M\}$
- **Detected Minutiae:** $F_d = \{f_d^1, f_d^2 \ldots, f_d^N\}$
- **Paired Minutiae (P):** if a detected minutia, f_d, lies within a distance threshold δ around a ground truth minutia f_g
- **Deleted Minutiae (D):** if a detected minutia, f_d, has to be deleted from the set F_d that could not be paired with any minutiae in the set F_g
- **Inserted Minutiae (I):** if a detected minutia has to be inserted to the set F_d to pair with



Evaluating Minutiae Detection

- Utilized Goodness Index (GI) [1] metric defined as:

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

where

- L = no. of 16 x 16 non-overlapping patches,
- Q_i = quality of the ith patch (good = 4, medium = 2, poor = 1),
- P_i = no. of paired minutiae in the ith patch,
- D_i = no. of deleted minutiae in the ith patch, such that ($D_i \leq 2 \times M_i$)
- I_i = no. of inserted minutiae in the ith patch, and
- M_i= no. of ground truth minutiae in the ith patch.

- Quality index by Chen et al. [2] is utilized to obtain Q_i

Larger the value of GI, better the performance of a minutiae extractor.

[1] Ratha, N. K.; Chen, S.; Jain, A. K.: Adaptive Flow Orientation-based Feature Extraction in Fingerprint Images. Pattern Recognition, 28(11):1657–1672, 1995.
[2] Chen, Y.; Dass, S. C.; Jain, A. K.: Fingerprint quality indices for predicting authentication performance. In: ICAVBPA. Springer, pp. 160–170, 2005.

Evaluating Minutiae Localization

- The positional error (e_p) for the paired minutiae ($\hat{f}_d, \hat{f}_g$) is computed using the Root Mean Square Deviation (RMSD) given by:

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

where (x_d^i, y_d^i) and (x_g^i, y_g^i) represent the locations of the detected and the ground truth minutiae, respectively.

- Similarly, the orientation error (e_θ) between the set of paired minutiae ($\hat{f}_d, \hat{f}_g$) is computed using:

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

where

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

Evaluate Robustness against Noise and Motion Blur

- *Stress testing* is used to test a system beyond normal operating conditions, often to a breaking point
- Noise model in Novetta's biosynthetic software is used to add controlled levels of (i) anatomical deformations (scars, holes, pressure variations), (ii) ridge noise, and (iii) finger dryness

Reference Fingerprint | **Noise Level 1** | **Noise Level 2** | **Noise Level 3** | **Noise Level 4**

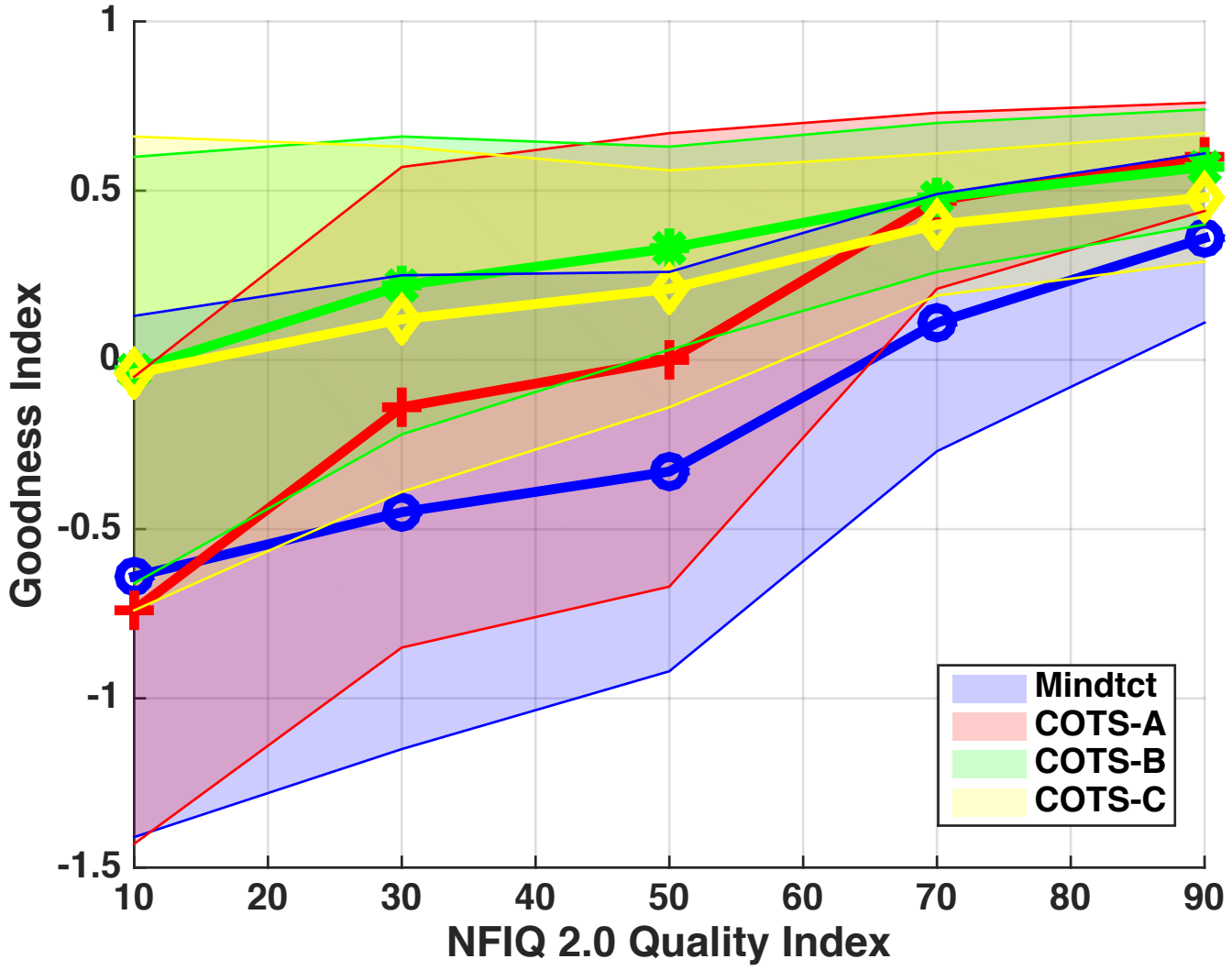
Evaluate Robustness against Noise and Motion Blur

- Movements of hand during fingerprint acquisition leads to motion blur
- Simulated three levels of motion blur using MATLAB's *motion lens filter* in both horizontal and vertical direction

Reference Fingerprint **Motion Blur Level 1** **Motion Blur Level 2** **Motion Blur Level 3**

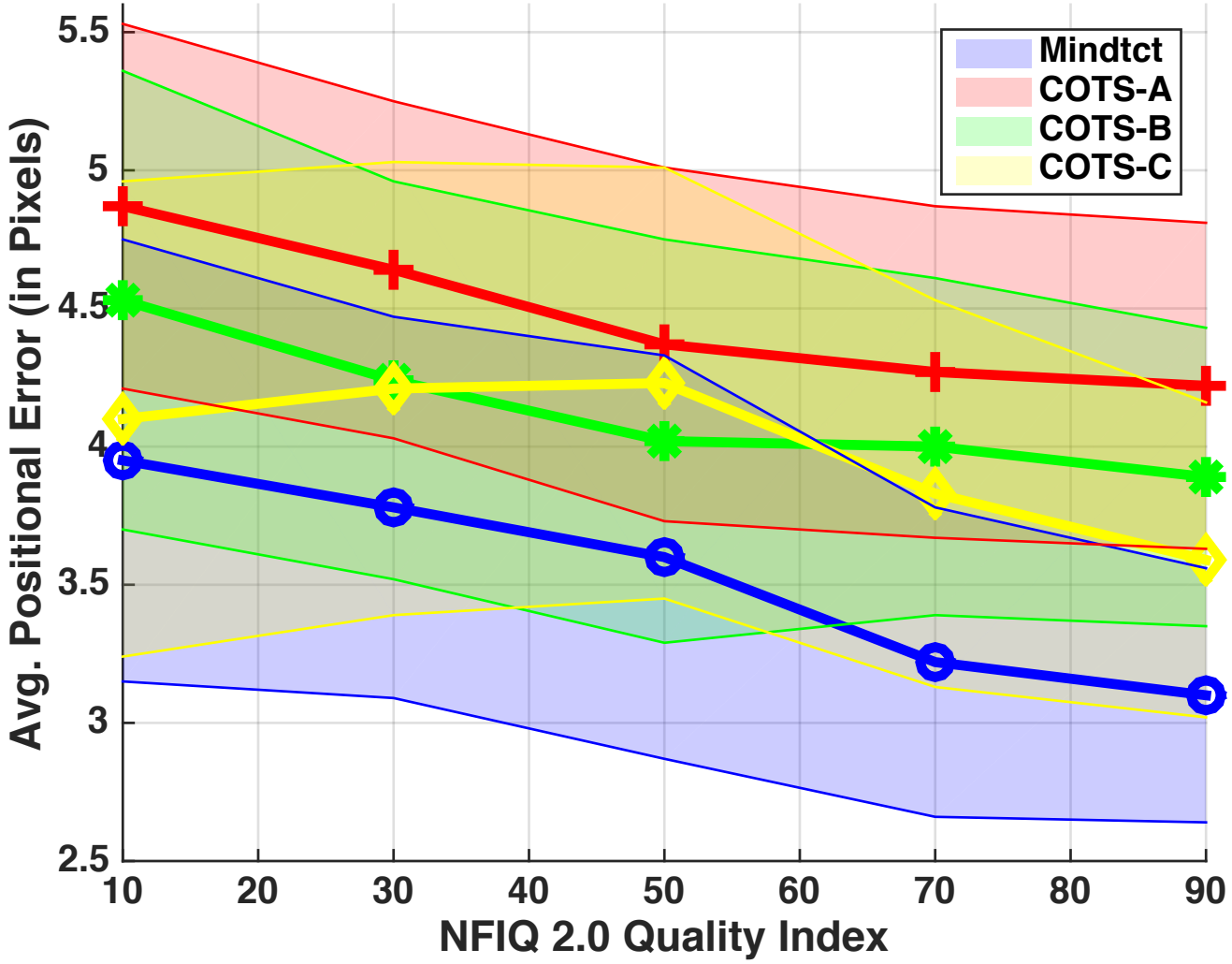
Experimental Results: Dataset-A Minutiae Detection

- Goodness Index
  - COTS-B and COTS-C exhibit high GI values
  - Mindtct and COTS-A produce large no. of spurious minutiae
  - Tipping point: NFIQ2 < 30



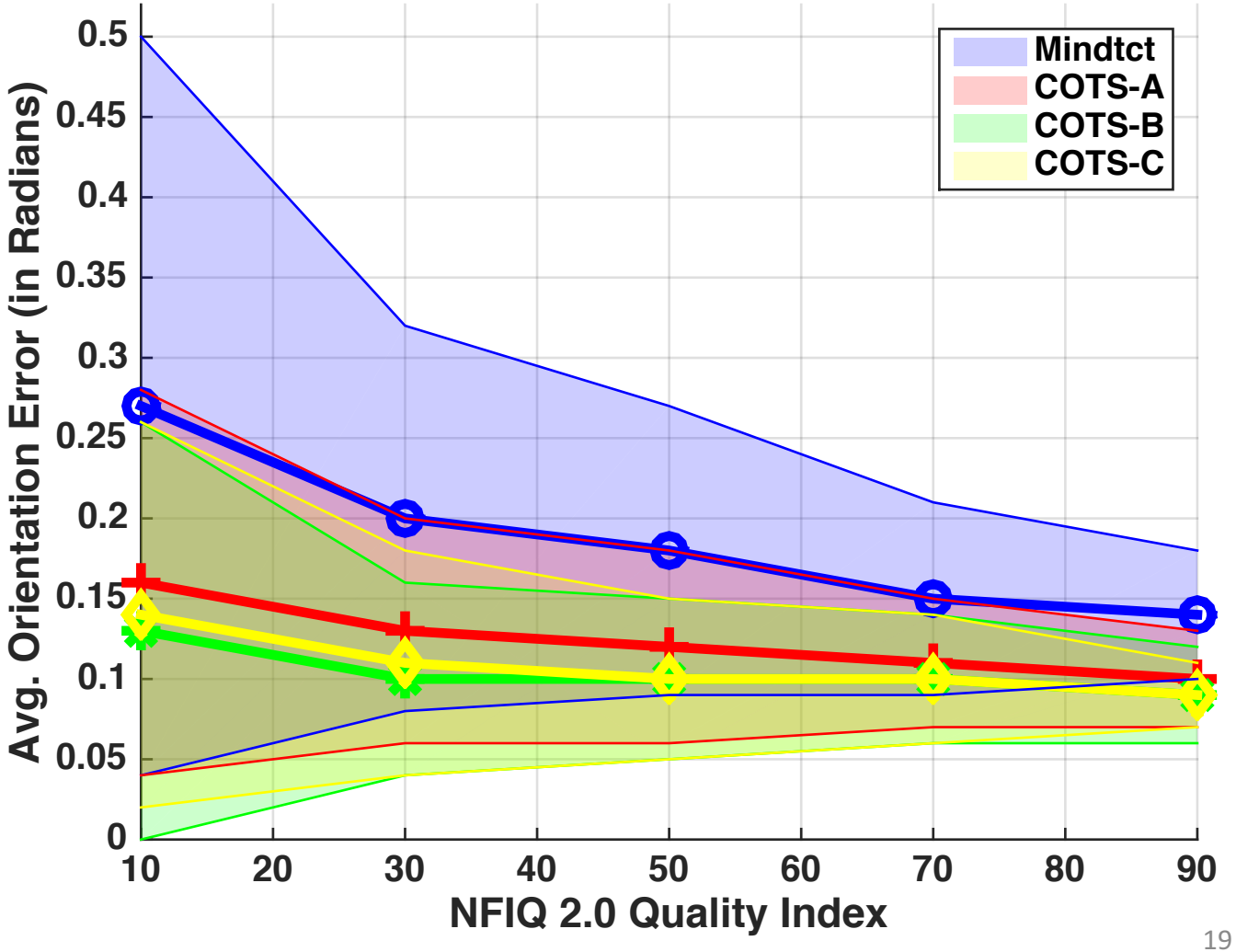
Experimental Results: Dataset-A Minutiae Localization

- Avg. Positional Error
  - Mindtct exhibits lowest positional error for paired minutiae



Experimental Results: Dataset-A Minutiae Localization

- Avg. Orientation Error
  - COTS extractors perform similarly, better than open-source extractor



Experimental Results: Dataset-A

Ground Truth
Manually Marked Minutiae

***mindtct* (open-source)**
Goodness Index : 0.47
Avg. Pos. Error (e_p): 2.65
Avg. Ori. Error (e_θ) : 0.13

COTS - A
Goodness Index : 0.90
Avg. Pos. Error (e_p): 4.41
Avg. Ori. Error (e_θ) : 0.07

COTS - B
Goodness Index : 0.77
Avg. Pos. Error (e_p): 4.68
Avg. Ori. Error (e_θ) : 0.05

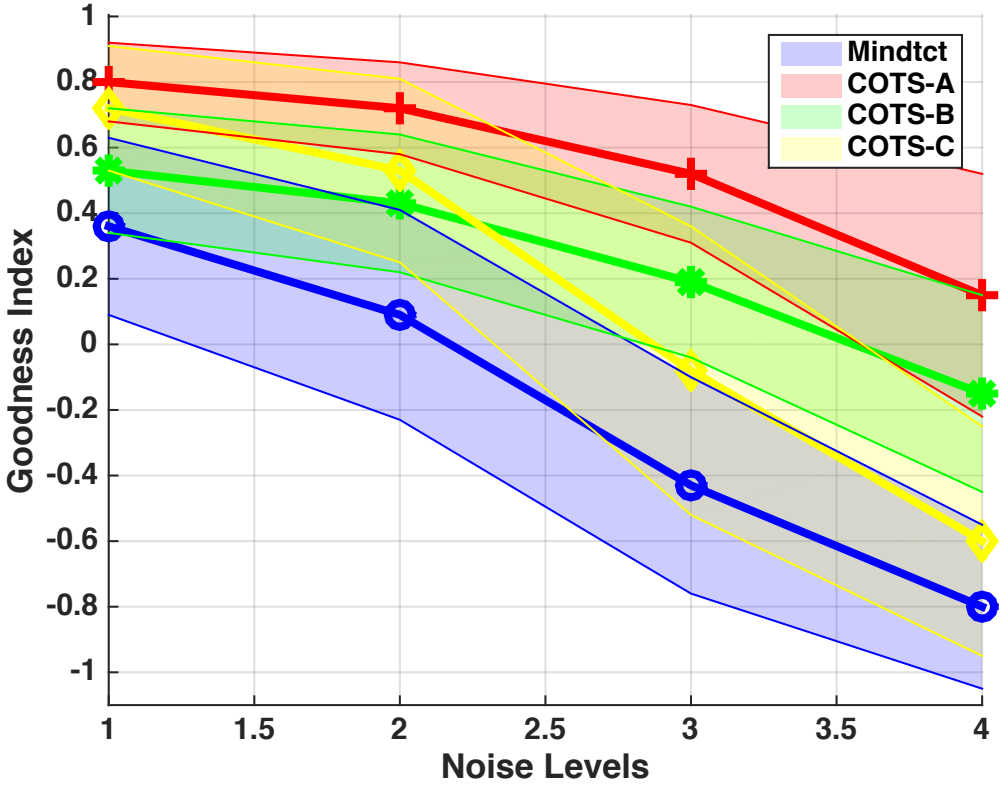
COTS - C
Goodness Index : 0.70
Avg. Pos. Error (e_p): 3.48
Avg. Ori. Error (e_θ) : 0.06

Example fingerprint from FVC2002 DB1A dataset. 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

Experimental Results: Dataset-B

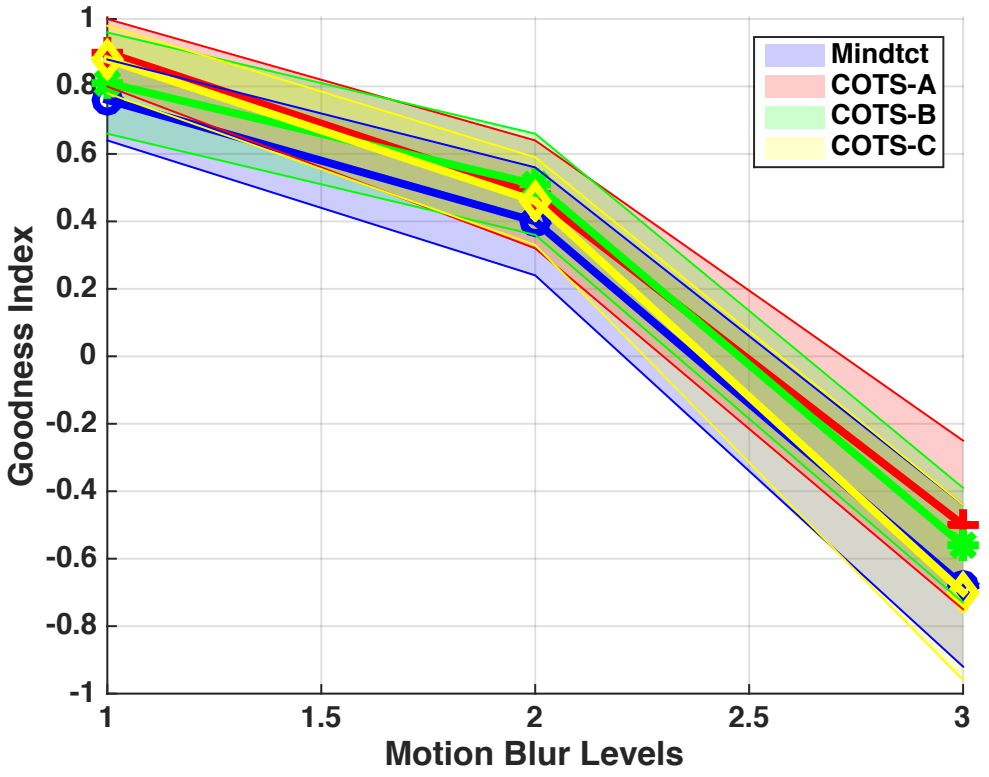
Goodness Index

Robustness against Noise



- COTS-A highly resilient to Noise
- Tipping Point: Noise level > 2

Robustness against Motion Blur

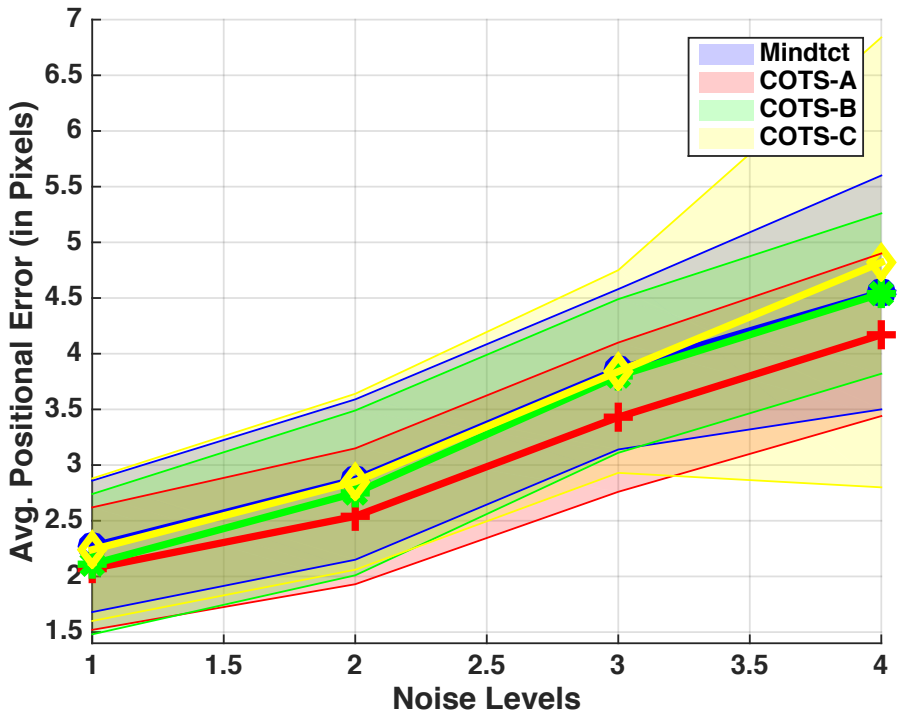


- Tipping Point: Motion Blur level > 2

Experimental Results: Dataset-B

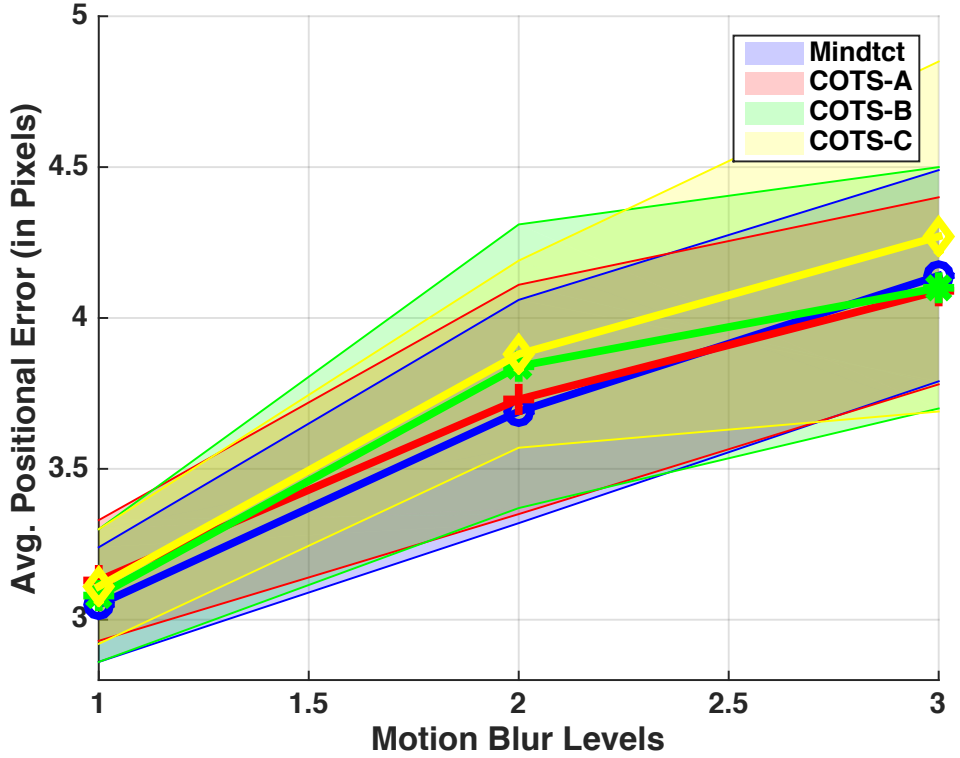
Avg. Positional Error

Robustness against Noise



- COTS-A exhibits lowest positional error
- High variance in COTS-C performance

Robustness against Motion Blur

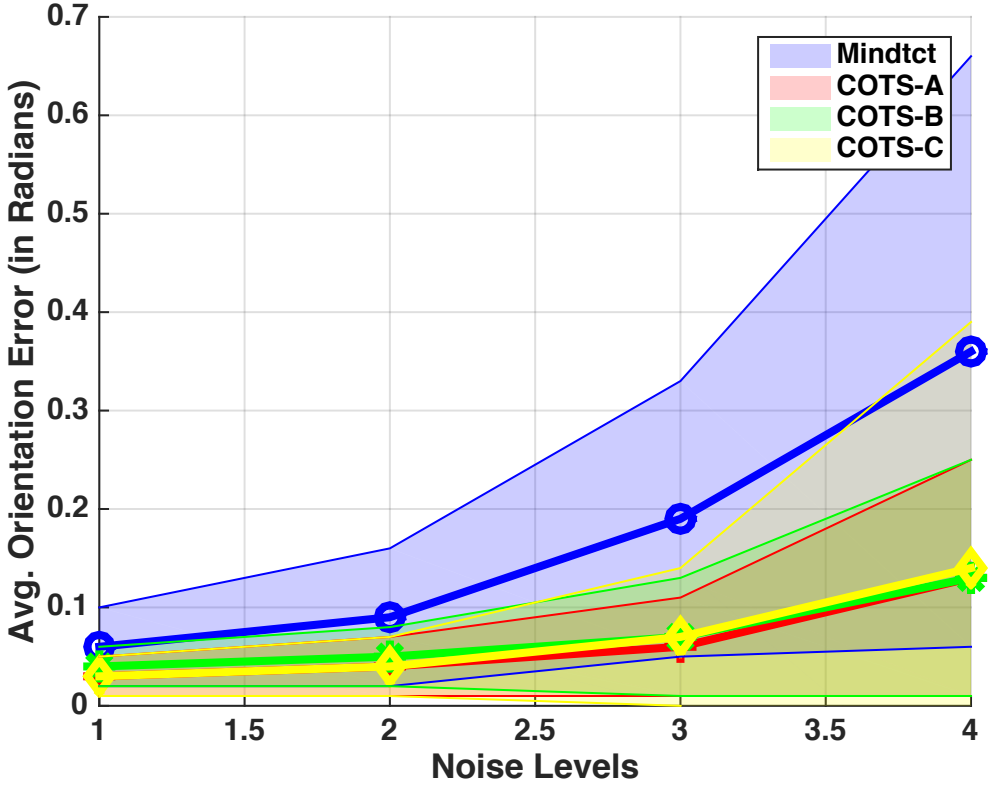


- COTS-A and open-source extractors exhibit lowest positional error
- High variance in COTS-C performance

Experimental Results: Dataset-B

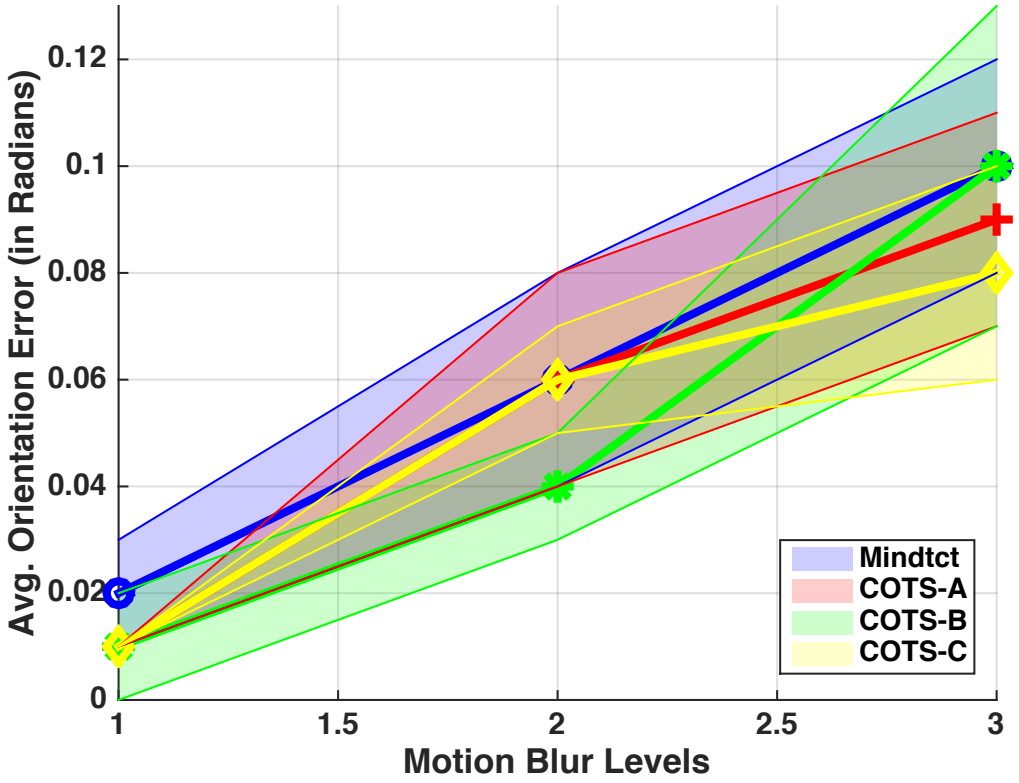
Avg. Orientation Error

Robustness against Noise



- Open-source extractor breaks above noise level 2 and exhibit high variance

Robustness against Motion Blur



- Lower orientation errors compared to noise degradations

Conclusions

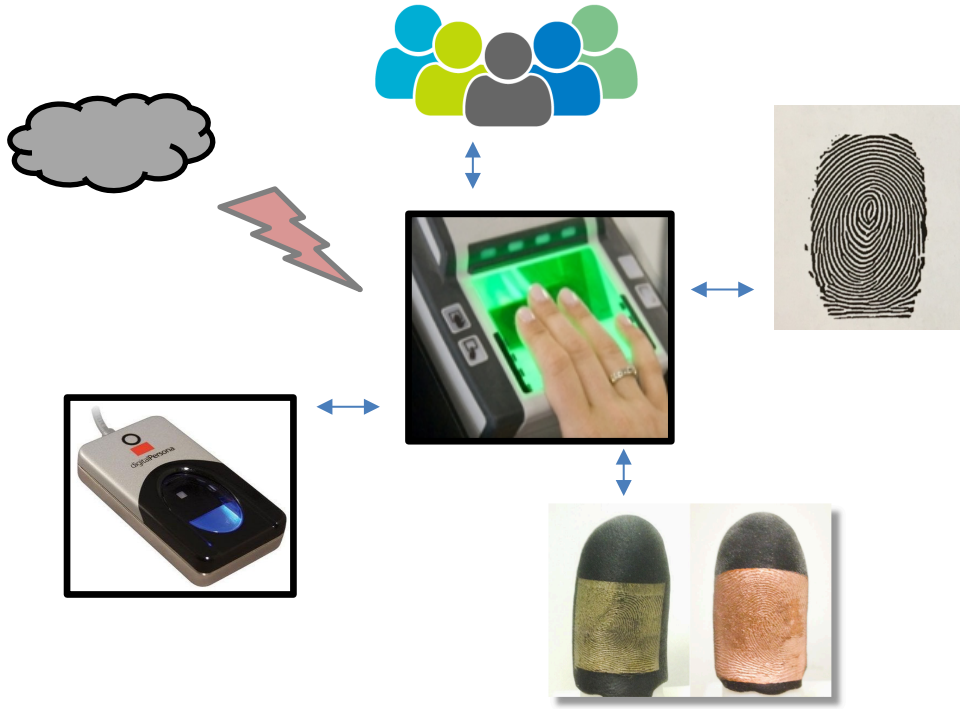
- Primary limitation of open-source minutiae extractor, mindtct, is spurious minutiae,
- On average, a 100% increase in positional error as Noise levels increase from 1 to 4,

Thank You

Any Questions?

Independent Evaluation of Fingerprint Sensors

- Fingerprint Certification Standards (PIV-071006 and Appendix F)
- Independent evaluations in literature
  - Resiliency to external environment (temperature and humidity)
  - Intrinsic subject-dependent factors (skin humidity and pressure)
  - Operational quality
  - Interoperability
  - Presentation attack detection
  - Calibration



Dataset-A

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)

Dataset-B (Noise)

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)

Dataset-B (Motion Blur)

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)