Face Recognition: Issues for DHS

Patrick Grother and Craig Watson
NIST
U. S. Department of Commerce

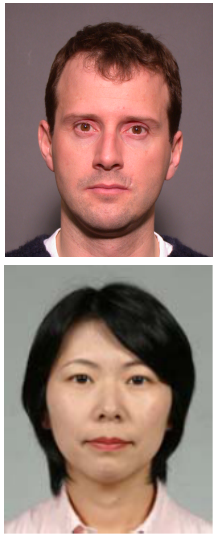
DHS Principals
March 27, 2020

The power of face recognition:

A demonstration

Step 1:
Enrol N = 104 million CBP ENTRY photos, of 32.6 million people

Source: FBI via NIST Special DB 32. Deceased subjects.



Step 2:
Search with almost ISO compliant frontals from DHS/CIS

Source: Author, ISO.

Accuracy: Two populations; two error rates

NIST

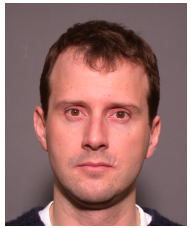
#1 » Enroll DHS/CBP ENTRY images

- 104.1 million
- 32.6 million people

#2 » Mated searches

- 2.3 million DHS/CIS APPLICATION images

→ "miss rate" aka False Negative ID rate



#3 » Non-mated searches

- 1.8 million DHS/CIS APPLICATION images

→ "false alarm rate" aka False Positive ID rate

104 Million: CIS to CBP search accuracy

INVESTIGATION FALSE POS ID RATE = 100%	NEC (0.7 + 1.1 seconds)	Rank One (0.1 + 18 seconds)
MISS RATE: Searches not returning ANY image of the correct person at rank 1	0.4%	2%

HIGH VOLUME, HIGH THRESHOLD IDENTIFICATION, FALSE POS ID RATE = 1%	NEC	Rank One
MISS RATE: Searches not returning ANY image of the correct person above threshold	0.6%	8.3%

Takeaways:

1. NEC was much better than anything else until recently:
   - ⇨ Sensetime (CN) 2020
2. But NEC version matters:
   - ⇨ Tokyo 2018 vs. TVS vs. OBIM vs. Tokyo 2020.
3. **Investigative search with N > 100M is possible, defensible**

Demographic effects in FR

NIST Interagency Report 8280

NISTIR 8280

Face Recognition Vendor Test (FRVT)

Part 3: Demographic Effects

Patrick Grother
Mei Ngan
Kayee Hanaoka
Information Access Division
Information Technology Laboratory

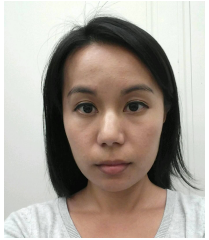
This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8280

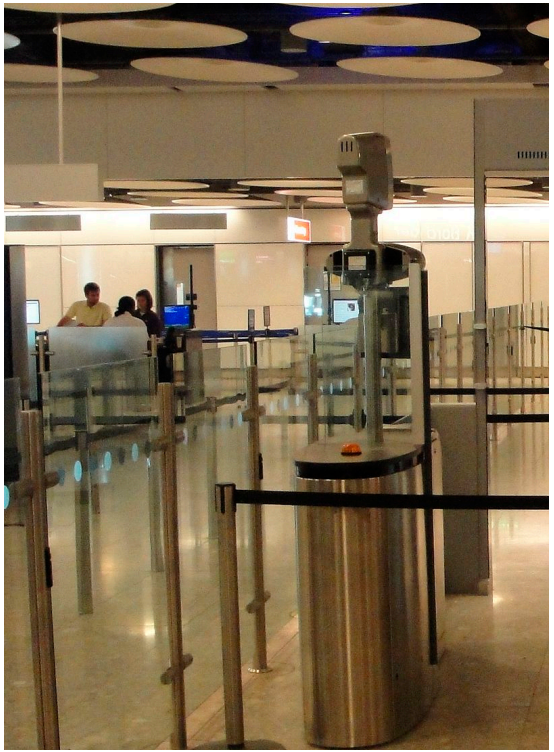
2019/12/19

National Institute of Standards and Technology
U.S. Department of Commerce

Inbound border crossing using passport verification

LIVE

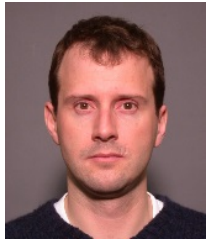




https://en.wikipedia.org/wiki/EPassport_gates
CC BY 2.0. File:Heathrow Terminal 5 ePassport gates.jpg
Created: 16 July 2010

CHIP

Wait for luggage



See an officer!

Two factor authentication:

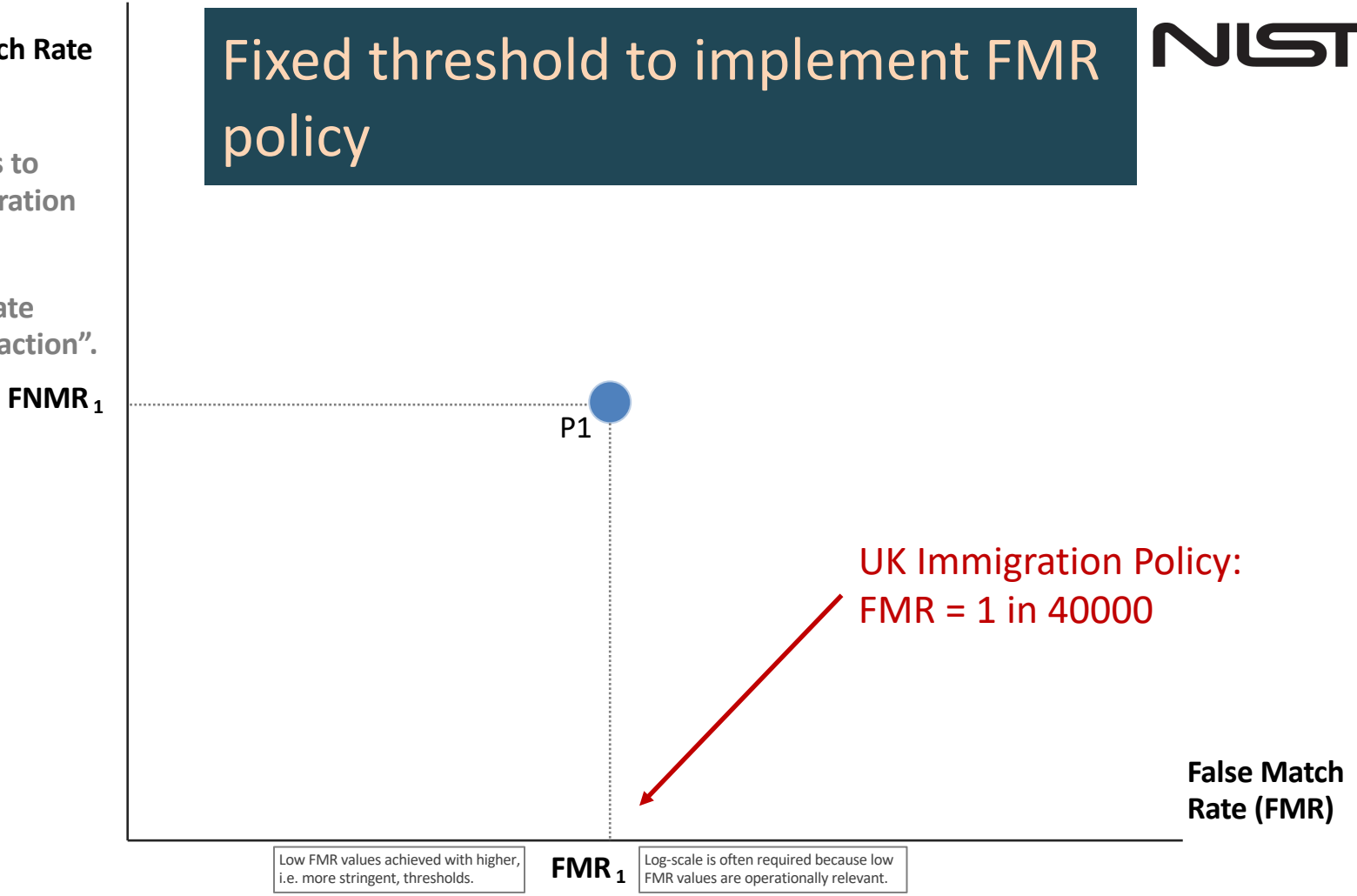
1. Something you have: Possession of passport
2. Something you are: Successful recognition of a biometric

Fixed threshold to implement FMR policy

False Non-Match Rate (FNMR)

Drives referrals to primary immigration officer

Related to e-Gate "traveler satisfaction".

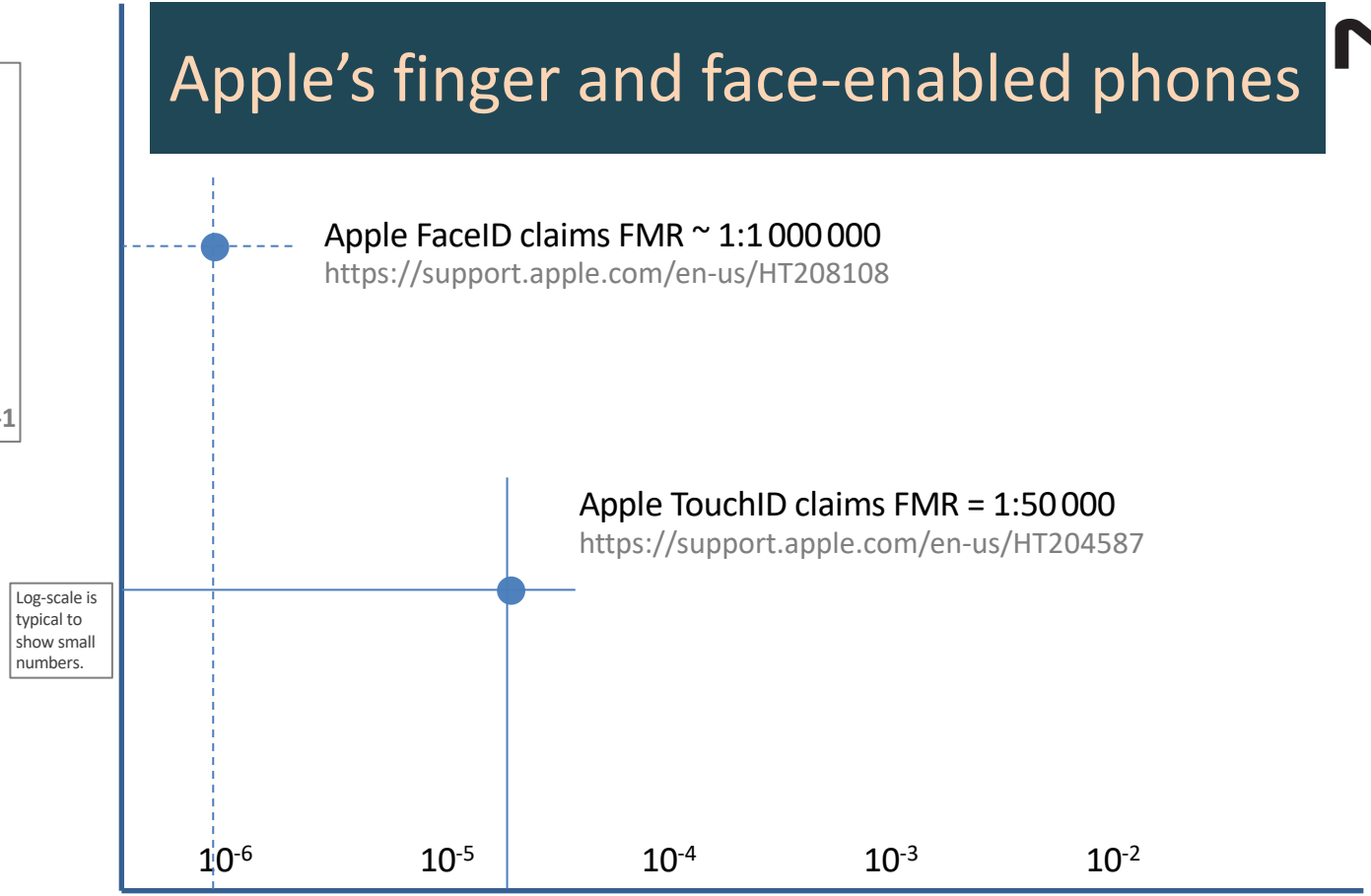


Apple's finger and face-enabled phones

FNMR
False non-match rate

Proportion of genuine comparisons producing score below threshold, T.

See ISO/IEC 19795-1



Log-scale is often required because low FMR values are operationally relevant.

FMR False match rate
Proportion of impostor comparisons searches yielding any candidates at or above threshold, T.

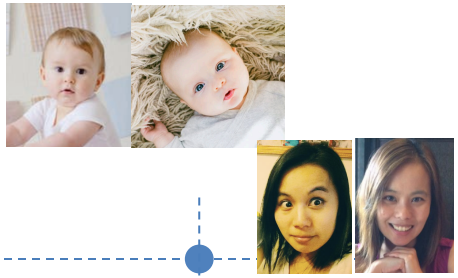
iPhoneX: Declared instance of demographic differential

FNMR
False non-match rate

Proportion of genuine comparisons producing score below threshold, T.

See ISO/IEC 19795-1

Log-scale is typical to show small numbers.



Apple Face. ID claims FMR ~ 1:1 000 000
https://support.apple.com/en-us/HT208108

"The statistical probability is different for twins and siblings that look like you and among children under the age of 13, because their distinct facial features may not have fully developed."

10^{-6} 10^{-5} 10^{-4} 10^{-3} 10^{-2} 10^{-1}

Log-scale is often required because low FMR values are operationally relevant.

FMR False match rate
Proportion of impostor comparisons searches yielding any candidates at or above threshold, T.

Detour

Twins: A Forgotten Demographic

Source: Twins Day Ohio collected by Notre Dame

Same person or not?

	Identical	Fraternal
How	Monozygotic	Dizygotic
Proportion of individuals that are a twin	0.9%	3.1%
Same-sex	100%	50% in theory 58% actually
TR gain since 1980	x1.5 since 1980	x1.9 since
Demographics	~ constant with age, geography	varies with mothers age, order, geography

Twins, triplets ... constituted 140,000 out of 4M births in 2015
https://www.cdc.gov/nchs/data/nvsr/nvsr66/nvsr66_01.pdf

Scenario: Identical Twins

Probe is an identical twin

Gallery Size: 1.6 million

Algorithm	Rank of sibling	Score	FPIR
Microsoft	1	0.778874	0.0007
NEC	1	0.768429	0.001
Idemia	1	3066.12	0.0007

Candidate List

Rank 1

...

...

Scenario: Siblings

Probe is a sibling

Gallery Size: 12 million

Algorithm	Rank of sibling	Score	FPIR
Microsoft	1	0.49	0.163
NEC	2	0.66	0.832
Idemia	11	525.6	0.997

Candidate List

Rank 1 (Microsoft) | Rank 2 (NEC) | … | Rank 11 (Idemia) | …

Age, sex, race effects

NIST Interagency Report 8280

NISTIR 8280

Face Recognition Vendor Test (FRVT)

Part 3: Demographic Effects

Patrick Grother
Mei Ngan
Kayee Hanaoka
Information Access Division
Information Technology Laboratory

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8280

2019/12/19

National Institute of Standards and Technology
U.S. Department of Commerce

Scope of NIST demographics work

- Algorithms
  - 187 algorithms, 99 developers
  - Mostly commercial, some universities
  - Prototypes from R&D labs
- Modes
  - One to one verification (DHS, DoS)
  - One to many identification (mugshots)
- Metrics:
  - False positives
  - False negatives
  - Failure to enroll
- Relevance to applications

- 18.3 million cooperative photos of 8.5 million people
  - DHS/CIS Application Photos
    - High quality
    - Race: 24 countries, 7 regions
    - Sex: M, F only
    - Age groups: [12-20], [20-35], [35-50], [50-65], [65-99].
  - DHS/CBP Entry Photos
    - Mediocre quality
    - Compare with excellent DHS/CIS photos
  - DOS Visa photos
    - Age
  - FBI mugshots
    - Sex: M, F, only
    - Age groups: Adults above or below 45.
    - Race: Asian, Black, White, Native American

Demographic effects, in brief

NIST

3. Algorithms

- Some much more accurate than others
  - Better accuracy ⟶ lower demographic diffs
- Some Chinese algorithms don't give high false positives in East Asian.
- A few 1:N identification algorithms "stabilize" false positive rates across demographics.
  - Including one from NEC
  - But ... versioning across DHS
- "KNOW YOUR ALGORITHM"
  - For demographic effects
  - For other image quality and other sensitivities

1. False positives

- AGE: Higher in the elderly and the very young
- SEX: Higher in women
- RACE:
  - Highest in East Asia, Africa
  - Lowest in Eastern Europe

FP >> FN: Security and civil liberties disparities

2. False negatives

- Somewhat higher in persons born in Africa
  - But error rates are low
- Somewhat higher in women
- Somewhat higher in children (ageing effects)
  - Not well characterized recently.

Mitigation

Near-term

- Training
  - Larger + more diverse sets
  - Cost functions (next slide)
- Improved awareness
  - Buy the correct algorithm
  - Operational tests “know your algorithm”
  - Lab tests.
- Closed loop feedback to developers
  - Need metrics
  - Need evaluations
  - Need image sets

Longer-term

- New signals
  - Skin texture?
  - Iris
- Better images
  - More standardized images
  - Face-aware capture devices
    - Exposure control
    - >8 bits capture
  - Higher resolution
    - For human adjudicators
    - For detection of active attack
    - For more discriminative information

Standards activities

ISO/IEC 24358
Face-aware capture

ISO/IEC 24357
Image quality assessment

ISO/IEC 19795-10
Performance testing for demographic effects

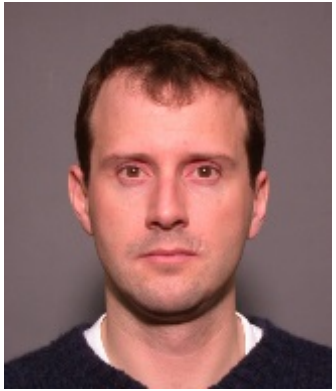
Twins can be distinguished:

Use of high-resolution photos

Source: Twins Day Ohio collected by WVU / Notre Dame

ISO/IEC 24358

FASTER, BETTER, FACE-AWARE CAPTURE (QUALITY MATTERS!)



Images from presenter

Problems:

a) Non-frontal faces

b) No-faces, multiple-faces

c) Over-, under-exposure

d) Human review errors

e) Morphing

f) Inadequate presentation attack detection

Potential Solutions:

a) Face pose detector

b) Face detectors

c) 12 bits or closed-loop control

d) Higher resolution, better compression, 3D

e) Crypto for tamper-proofing

ISO/IEC 24358 Face-aware Capture: Est. pub 2022

NIST

OBJECTIVES	CANDIDATE TECHNOLOGIES TO MITIGATE
1. Reduce photos with 0 people, > 1 person	» In-the-loop face detector
2. Reduce pose-variation	» In-the-loop pose estimation
3. Reduce under- and over-exposure	» 10 bit sensors; comp. photography
4. Improve PAD	» More resolution
5. Improve MAD	» More resolution
6. Improve human adjudication	» More resolution, more views
7. Eliminate tampering	» Crypto, watermarking
8. Support disambiguation of (identical) twins	» More resolution – provable to receiving system

ISO/IEC 24357 Face-image quality: Est. pub 2022

NIST

Objectives

- Improve face image quality
  - Reduce false negatives
  - Reduce false positives
- Improve conformance to ISO face photo specifications
  - ISO/IEC 39794-5

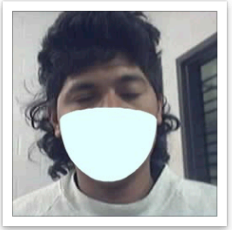
Supporting test program

- FRVT quality
- Check whether image quality values predict recognition failure
- Initially: DHS/CBP and DHS/CIS images

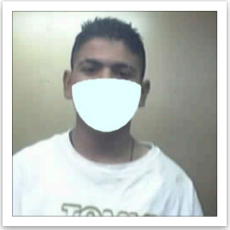
Standard requirements

- Software must measure and report values for
  - head pose
  - resolution
  - presence of one face
  - occlusion
  - over- and under-illumination
  - eyes-closed
  - …
- Software must
  - Provide actionable feedback
  - summarize quality
  - select “best” image
  - aggregate quality values (e.g. across dates, sites, populations)

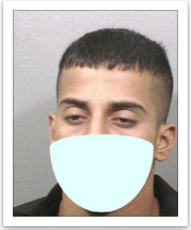
Effect of face masks on FR



S050-01-t10_01.ppm.jpg



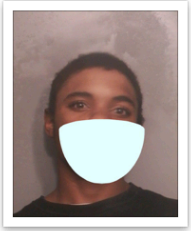
S051-02-t10_01.ppm.jpg



S052-02-t10_01.ppm.jpg

S058-02-t10_01.ppm.jpg

S059-01-t10_01.ppm.jpg



S060-01-t10_01.ppm.jpg

- Synthetic face masks
- Apply to DHS images: Estimate effect on face accuracy.
- Status:
  - Initial look: Promising
  - Formal accuracy estimates on-hold pending access to NIST servers.

THANKS

PATRICK.GROTHER@NIST.GOV

FRVT and Beyond

CURRENT

1. FRVT 1:1
Core Biometric Operation

Report: 2020-03-25

2. FRVT 1:N
Search Performance

Report: 2020-03-24

3. FRVT Morph
Detection

Report: 2020-03-04

4. FRVT Q
Automated Quality Assessment

Report: 2020-03-06

7. NOT FRVT
Characterizing Human FR Capability

PAST

5. FIVE
Recognition in video

Report: 2017-05

6. FRVT AGE / SEX
Estimation from single image

Report: 2014-03

FUTURE (?)

8. FRVT PAD
Presentation Attack Detection

9. Template Protection Evaluation

10. De-Identification for Privacy

11. Age Estimation