AGENDA

TOPIC	120 MINUTES	PRESENTER
NIST Introduction	**5 minutes**	**Craig**
Biometrics 101	**10 minutes**	**Craig**
Face	**52 minutes**	
Face: Introduction	4	Patrick
Face: 1:1 State of the Art	6	Patrick
Face: 1:N State of the Art	6	Patrick
Face: Ageing	3	Patrick
Face: Demographics	8	Patrick
Face: Twins	2	Patrick
Face: Human Role	6	Patrick
Face: Morph Attack	8	Mei
Face: Presentation Attack	8	Mei
Iris	**8 minutes**	**Patrick**
Q&A	**10 minutes**	**Moderator: Craig**
AEV	**8 minutes**	**Mei**
Contactless Fingerprint	**10 minutes**	**Craig**
Human/Device Interaction	**10 minutes**	**Craig**
Q&A	**5 minutes**	**Moderator: Patrick or Mei**
Wrap-up Summary	**2 minutes**	**Patrick**

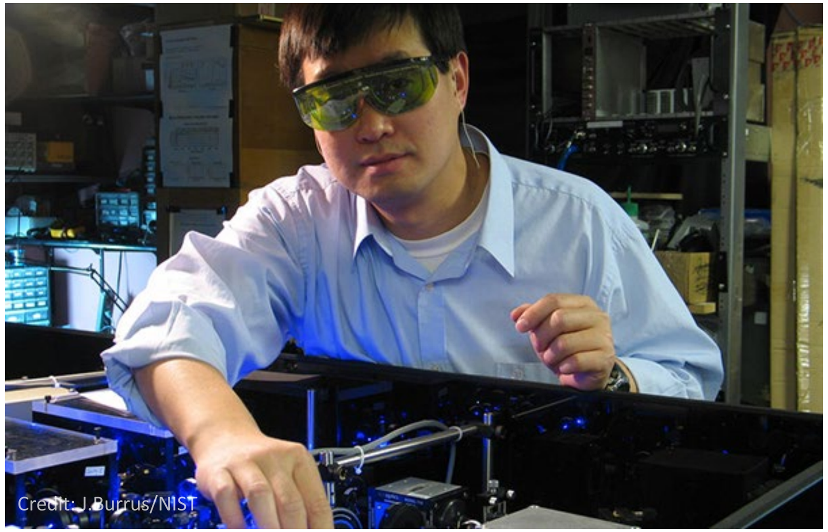
CRAIG WATSON PATRICK GROTHER MEI NGAN

Slides are available – link on the last slide

NIST: A Brief Introduction

NIST Mission

To promote U.S. innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve our quality of life



NIST at a Glance

NIST

3,400+
FEDERAL EMPLOYEES

5
NOBEL PRIZES

2 CAMPUSES
GAITHERSBURG, MD [HQ]
BOULDER, CO

3,500+
ASSOCIATES

10
COLLABORATIVE INSTITUTES

400+
BUSINESSES USING NIST FACILITIES

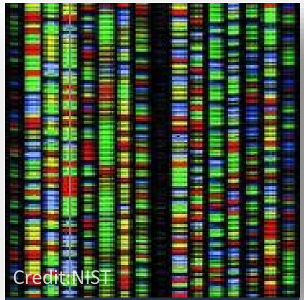
ManufacturingUSA

16
NATL OFFICE FOR MANUFACTURING INSTITUTES

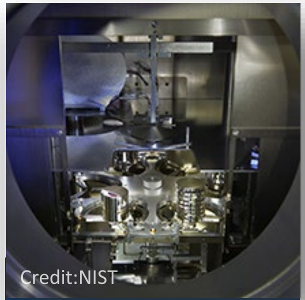
51
MANUFACTURING EXTENSION PARTNERSHIP CENTERS

U.S. BALDRIGE PERFORMANCE EXCELLENCE PROGRAM

NIST Laboratory Programs



Material Measurement Laboratory



Physical Measurement Laboratory

Engineering Laboratory

Information Technology Laboratory

Communication Technology Laboratory

NIST Center for Neutron Research

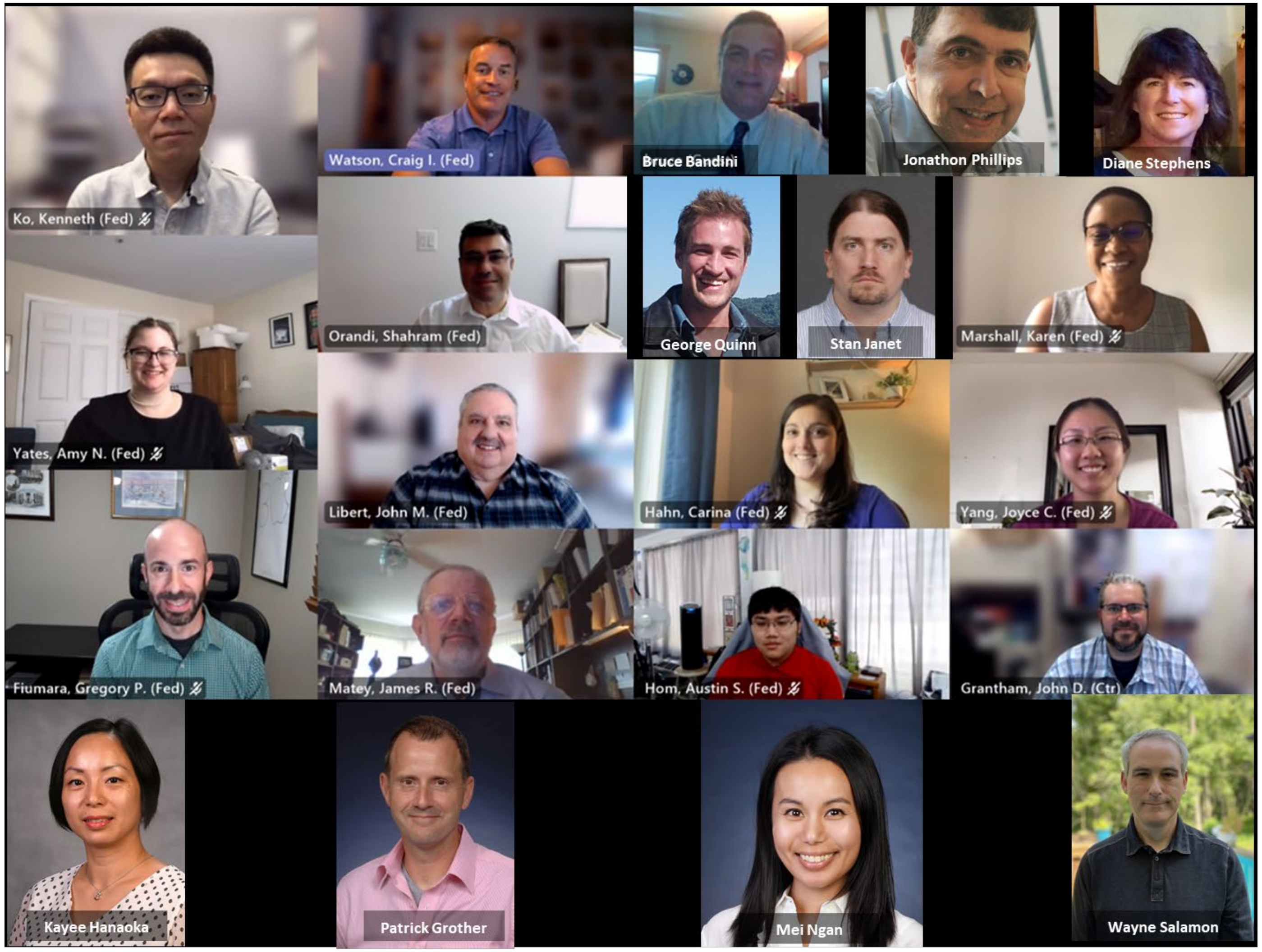
Information Access Division: Image Group

NIST

Information Technology Lab → **Information Access Division** → **Image Group**

Ensuring identity for trust in commerce and justice

NIST's Biggest Strength: Our Reputation

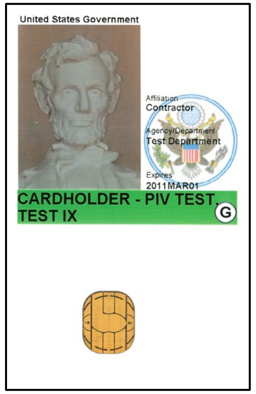


- Technical excellence
- Integrity
- Uncompromising
- Rigorous
- Unbiased
- Industry focused
- Non-regulatory

Interoperability: “Common” Language

Commerce: Digital Identity and Fraud Prevention

https://www.idemia.com/walk-through-multi-biometric-solution



https://www.irisid.com/productssolutions/hardwareproducts/icam-d2000/

https://tascent.com/wp-content/uploads/2022/03/tascent-insight-one-self-service-solutions-brochure-2018.pdf

Automated Border Control Gate

Source: http://www.futuretravelexperience.com/2016/01/automated-border-control-e-gates-go-live-at-naples-airport/

Source: Dulles CBP's New Biometric Verification Technology Catches Third Impostor in 40 Days | U.S. Customs and Border Protection

CBP Simplified Arrival

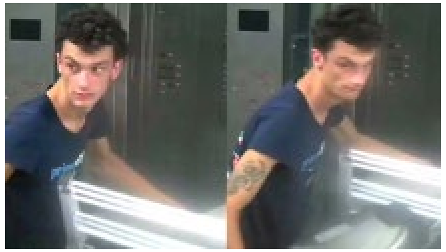
Source: https://www.cbp.gov/newsroom/local-media-release/cbp-introduces-simplified-arrival-denver-international-airport

Justice: Accurate Identification

Jarrod W. Ramos
Credit.. Anne Arundel Police, via Associated Press

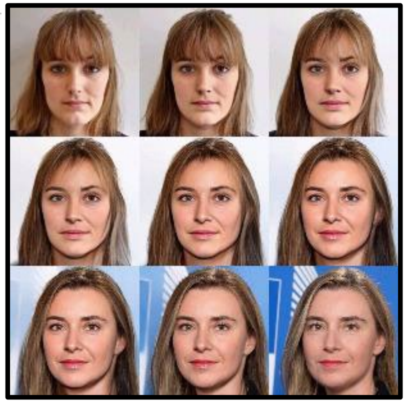
Sourrce: Facial recognition technology used in murder arrest | Las Vegas Review-Journal (reviewjournal.com)

Sourrce: NYPD uses facial recognition to arrest brazen sex offender accused of attempted rape on subway platform | Fox News



Sourrce: Man Charged After 3 Rice Cookers in Manhattan Spark Rush-Hour Scare – NBC New York

Source How Facial Recognition Is Fighting Child Sex Trafficking | WIRED



Source (9/22/2018): http://www.spiegel.de/netzwelt/netzpolitik/biometrie-im-reisepass-peng-kollektiv-schmuggelt-fotomontage-in-ausweis-a-1229418.html via Google Translate

Trustworthy & Fair: Understanding Capabilities and Limitations

https://www.cbsnews.com/news/facial-recognition-60-minutes-2021-05-16/

INCORRECT PERSON
NIJEER PARKS

https://www.cnn.com/2021/04/29/tech/nijeer-parks-facial-recognition-police-arrest/index.html

https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html

Biometrics 101

"the measurement and analysis of unique physical or behavioral characteristics (such as fingerprint, face, iris, or voice patterns) especially as a means of verifying personal identity"

Source: https://www.merriam-webster.com/dictionary/biometrics

Desirable Traits...

of a biometric

Universality - *we all have it*

Uniqueness - *distinguishing*

Permanence - *stable over time*

Measurability - *can be sensed*

Acceptability - *ease of use*

Circumvention - *no spoofing*

Performance - *accurate*

Jain A.K., Bolle R., and Pankanti, S., *Biometrics: Personal Identification in Networked Society,*
Norwell, Mass.: Kluwer Academic Publisher (1999).

of an algorithm

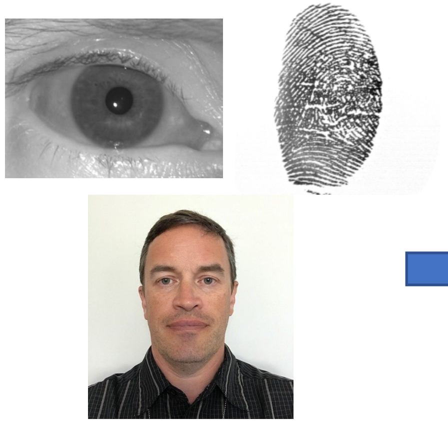
- Error rates (FMR, FNMR) are small
- Error rates (FNIR, FPIR) are low in large populations
- Accuracy – template size tradeoff exists
- Accuracy – speed tradeoff exists
- Memory requirements low and understood
- Error rates (FMR, FNMR) same across demographics
- FMR is stable under changes of the data
- Non-reversible templates
- …

Biometric Modalities

Face	• NIST FRTE/FATE (formerly FRVT) 1:1, 1:N, Video, Sex, Age, Quality, Pose Estimation
Fingerprint	• NIST FpVTE, MINEX, PFT, ELFT
Iris	• NIST Iris Exchange (IREX)
Voice	• NIST Speaker Recognition Evaluation
DNA	• NIST Advanced Chemistry Laboratory
Behavioral	• Gait, gesture, keystroke dynamics

…

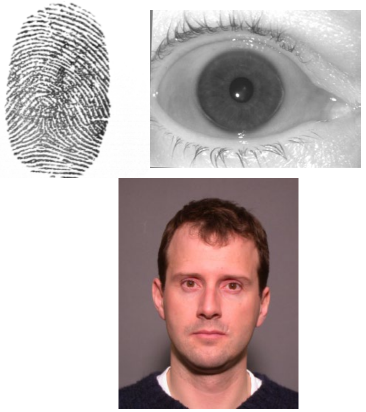
Use case: One-to-one (1:1) Verification



TWO IMAGES

(Identity Extractor)

(Identity Extractor)



Template

Template

Template Comparator

Match Score

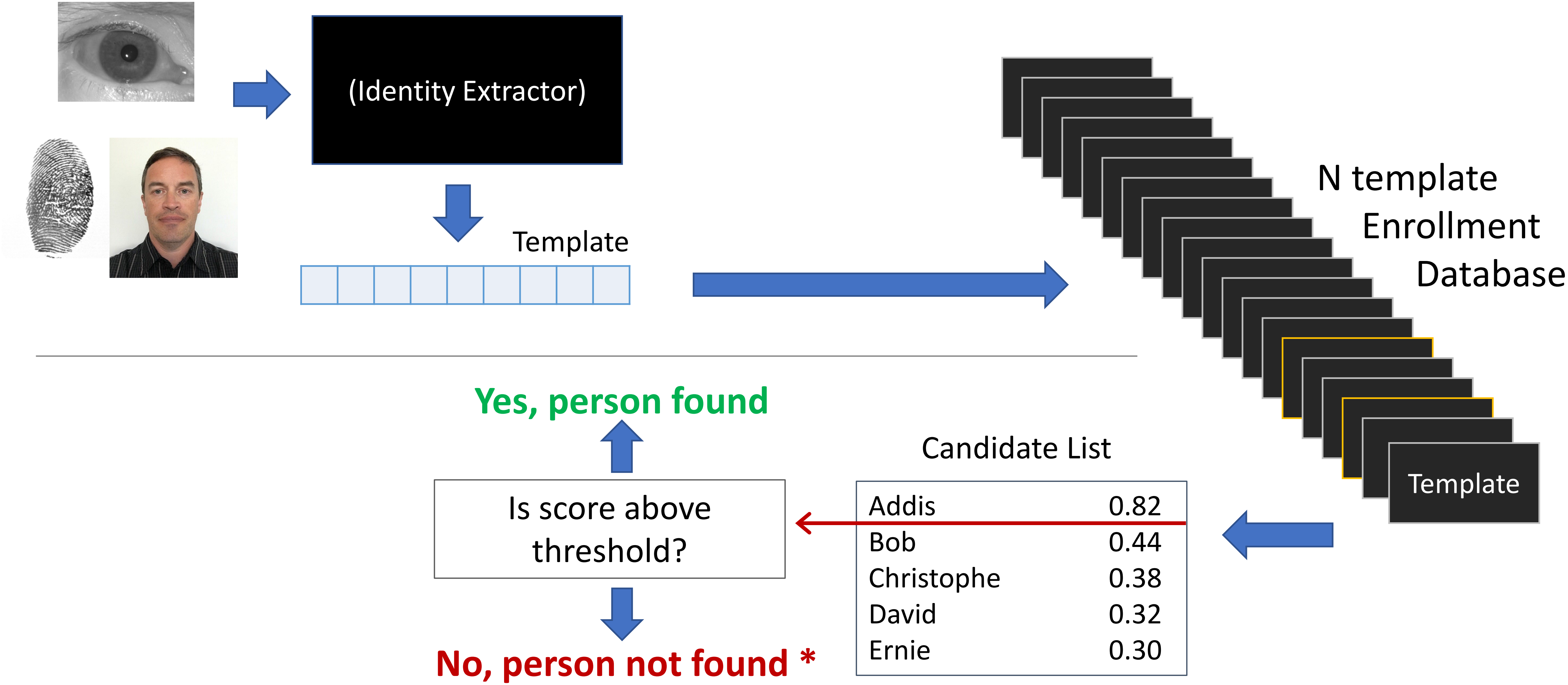
Is score above threshold?

No, different person*

Yes, same person!

Threshold is set by system owner to implement a policy

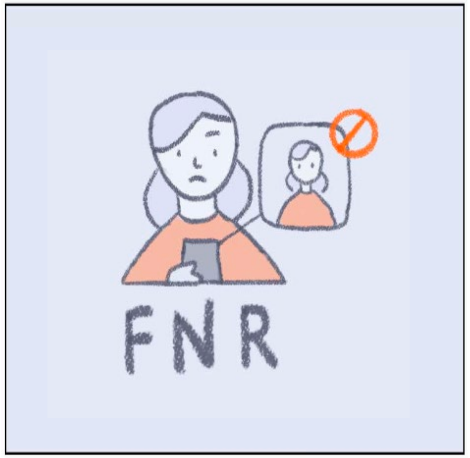
Use case: One-to-many (1:N) Identification



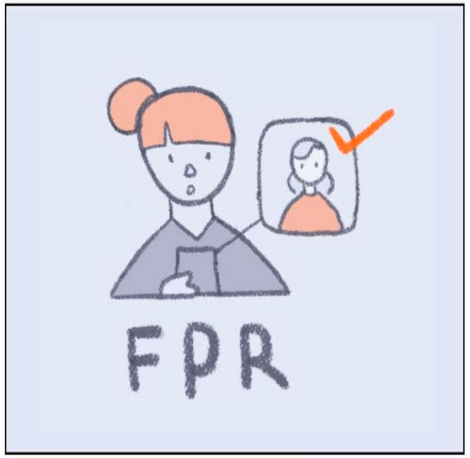
Threshold is set by system owner to implement a policy

Measuring Core Biometric Accuracy

Measuring Core Biometric Accuracy

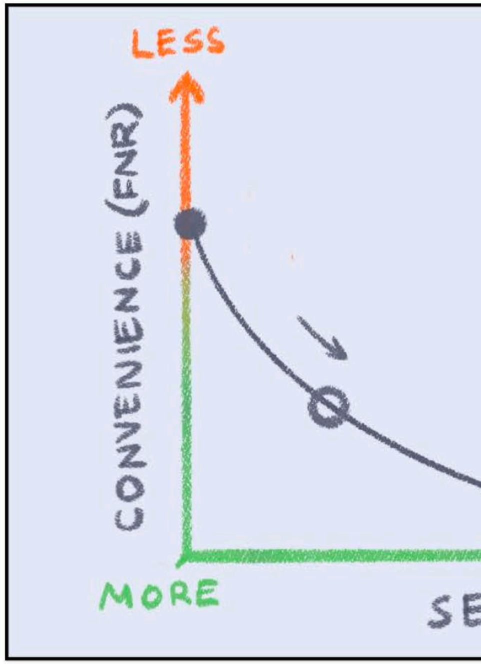


False Negative Rate is the rate at which a system fails to correctly match two samples of one person



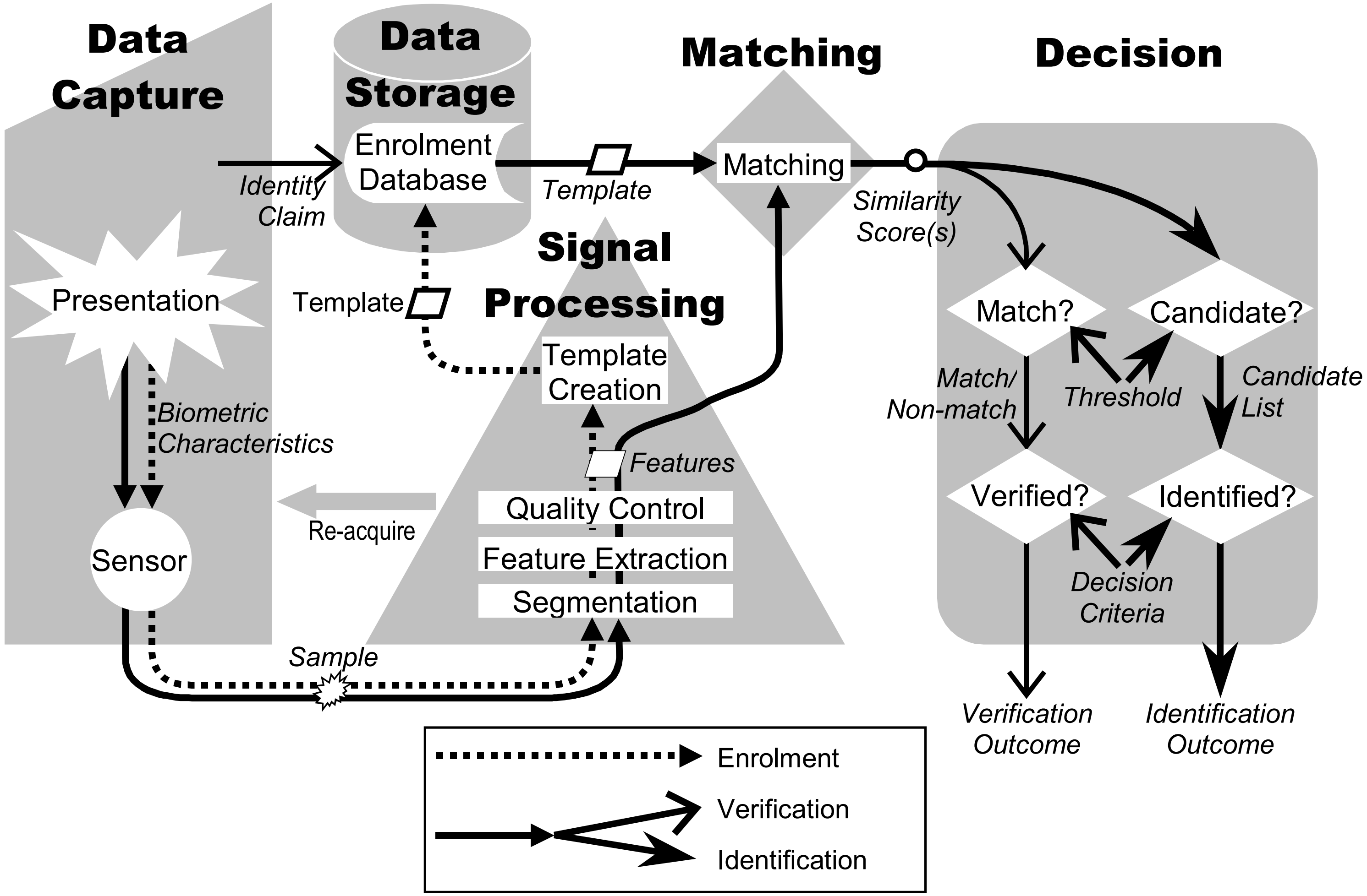
False positive rate is the rate at which a system incorrectly matches samples of two people

The proportion of false positives and false negatives determines the accuracy.



Depending on how high or low your False Positive Rate target is, you either get less security for more convenience, or more security for less convenience

SC37's Generic Biometric System



Source: ISO/IEC 19795:2006 — Information Technology — Biometric Performance Testing and Reporting — Part 1: Principles and Framework. This figure occurs in other SC37 standards also.

More metrics…

NIST

Capture or Processing

- FTE – Failure to enroll an individual
  - Usually transactional result
- FTA – Failure to acquire
  - Recognition-phase analog of FTE
  - Or a.k.a. FTS – failure to sense
- FTX – Failure to extract features
  - Software failure to make template
- FTP – Failure to process
  - Non-specific failure

1:1 Verification

- Matching (of templates)
  - FMR – False Match Rate
  - FNMR – False Non-match Rate
- Transactional results
  - FAR – False Accept Rate
  - FRR – False Reject Rate

Attack Detection

- BPCER
  - False Detection Rate
- APCER
  - Missed Attack Rate
- IAPMR
  - Attack + Match Success Rate

1:N Identification

- Matching
  - FPIR – False Positive Identification Rate
  - FNIR – False Negative Identification Rate
- Transactional
  - FPIR
  - FNIR
- In AFIS law enforcement
  - Reliability, Hit / Miss
  - Selectivity, False Alarm

1:N Investigation

- CMC
  - Rank based metric

The wide world of biometric testing

NIST

1. Technology Testing:

- Usually offline, with images
- Usually algorithms, could be cameras
- Why?
  - Scales to large size
  - Repeatable, so fair
  - Comparative testing
  - Inexpensive
- Why not?
  - Doesn't measure camera rejections, if any
  - Doesn't measure post-recognition human involvement

2. Scenario Testing:

- Human-in-the-loop
- Representative volunteer population
- In a purpose-built environment mimicking an operation, "in vitro"
- Why?
  - To answer camera-human interaction questions
  - To manipulate environmental factors
- Why not?
  - Not exactly repeatable
  - Population limited to hundreds by time, cost

3. Operational Testing:

- Human-in-the-loop
- Operational population
- In the operational environment, "in vivo"
- Why?
  - To answer questions about the actual performance
- Why not
  - Requires instrumentation of the actual system
  - You don't know who is an impostor, and may not find out

Qualifying "our algorithm is 99.5% accurate"…

Accuracy + Resources

- Is it a rank-based hit rate?
  - What rank?
  - How big is the database?
- Is it a true accept rate?
  - At what false accept rate?
- Resources
  - Speed
  - Size

Use Case Considerations

- Risk
  - How likely is an imposter?
  - How likely is an attack?
- Impact
  - What is the impact of a false positive?
  - What is the impact of a false negative?
  - How to resolve failures?
- Balanced Performance Across All Users
- Age / Race / Sex

Face Recognition & Face Analysis

A. STATE OF THE ART
B. TWINS
C. AGEING
D. SEARCH
E. HUMAN ROLE + CAPABILITY + TRAINING
F. DEMOGRAPHICS
H. MORPH ATTACKS
I. PRESENTATION ATTACKS

NIST FACE BENCHMARKS

NIST

FRTE

FACE RECOGNITION TECHNOLOGY EVALUATION

RECOGNITION: **WHO** IS IN AN IMAGE

1:1 VERIFICATION	Same person or not?
1:N SEARCH	Who? Where? When?
TWINS DISAMBIGUATION	Same person, or twin?
FACE IN VIDEO 2024	People on the move

FATE

FACE ANALYSIS TECHNOLOGY EVALUATION

ANALYSIS: **WHAT** ABOUT AN IMAGE

MORPH DETECTION	Two people in one photo?
QUALITY SUMMARIZATION	Will this photo match?
QUALITY DEFECT DETECTION	How is this photo bad?
PAD	Subversive photo?
AGE ESTIMATION	How old? Old enough?

Benchmarks are:

- Independent
- Free
- Regular
- Fast
- Repeatable
- Fair
- Black box
- IP-protecting
- Open globally
- Large-scale
- Sequestered datasets
- Statistically robust
- Public
- Transparent
- Extensible
- **ABSOLUTE ACCU**
- **RELATIVE ACCU**

QUESTION :: HOW ACCURATE IS FACE RECOGNITION?

ANSWER: Face search will succeed 100% of the time if

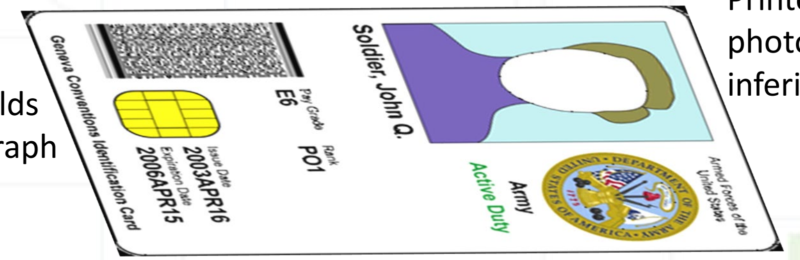
a. You're using a recent leading FR algorithm AND
b. There's a mate in the database AND
c. The mate is not more than X years old AND
d. The image is not manipulated AND
e. The image has limited quality problems – within the "capture envelope"

but there are caveats: Twins, attacks, demographics, application details

GOOD → BAD → UGLY

THE STATE OF THE ART :: 1:1

Chip holds photograph



Printed photograph inferior quality

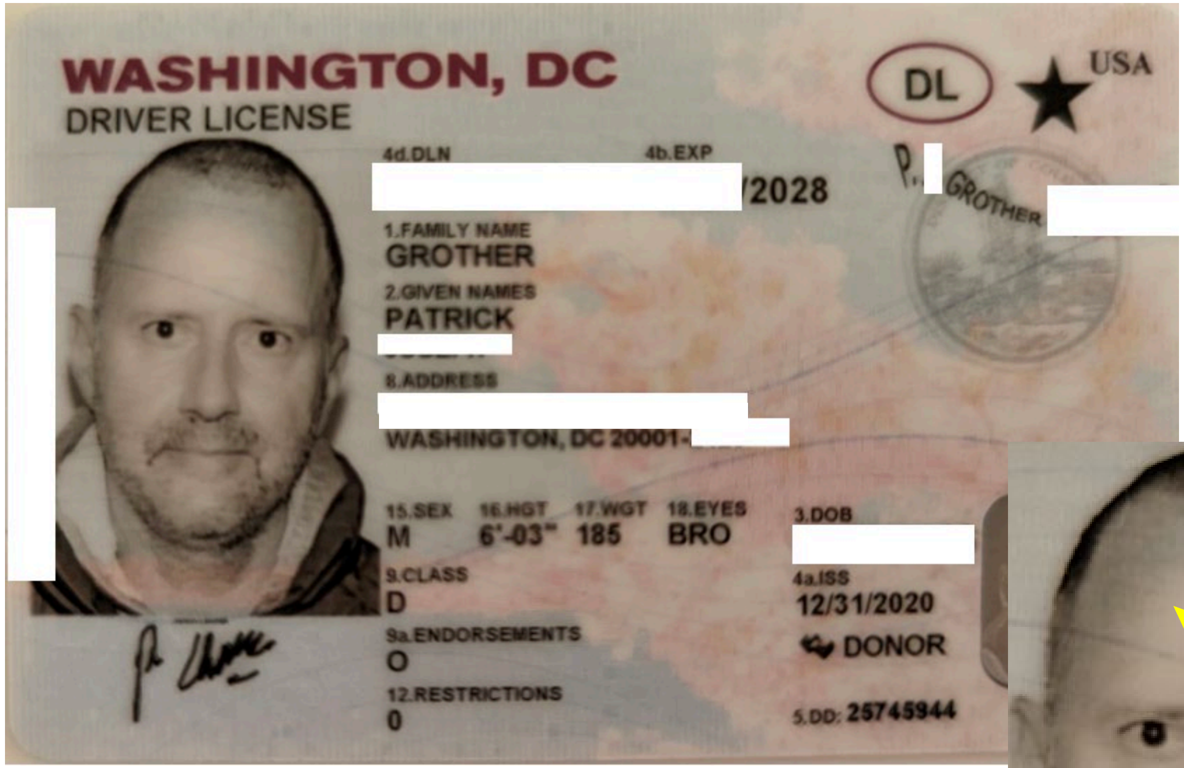
https://www.cnbc.com/2017/11/02/iphone-x-shipping-ahead-of-schedule-for-some-people.html

e-Passport + other ID credentials

https://images.app.goo.gl/8h3KAtn4mdJSvVuG8

USING FR TO BIND LIVE-PERSON TO ID CREDENTIAL

NIST



1. Scan ID credential

2. Segment

3. Live Image from phone or tablet

Face Recognition Verification

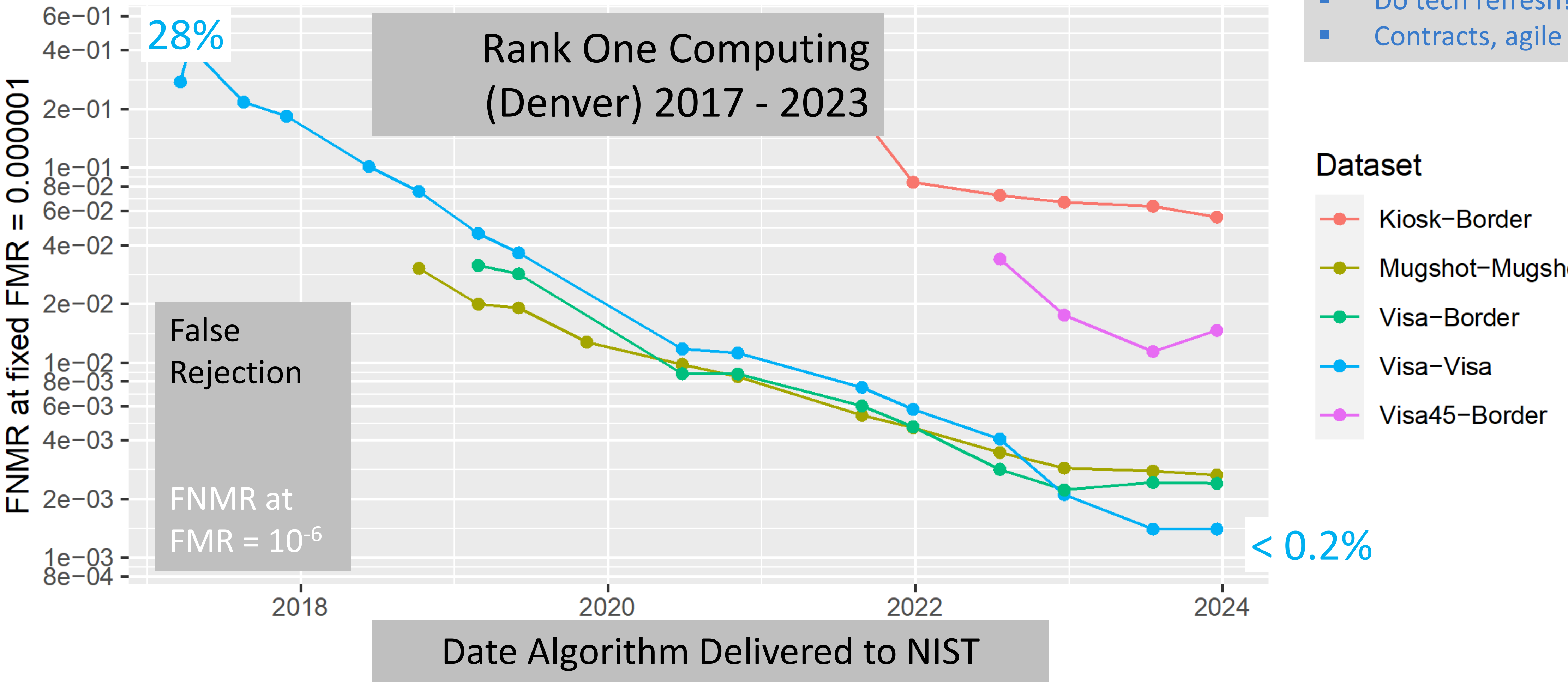
4. Same person, or not?

ACCURACY GAINS CONTINUE

NIST

Implications:

- Algorithms improve regularly
- Do tech refresh!
- Contracts, agile procurement

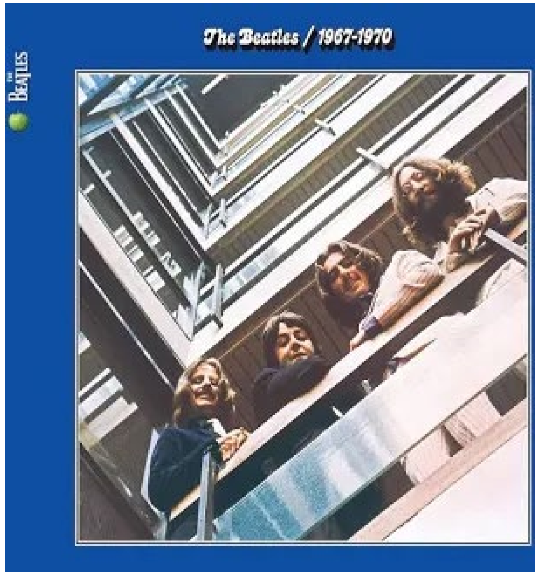


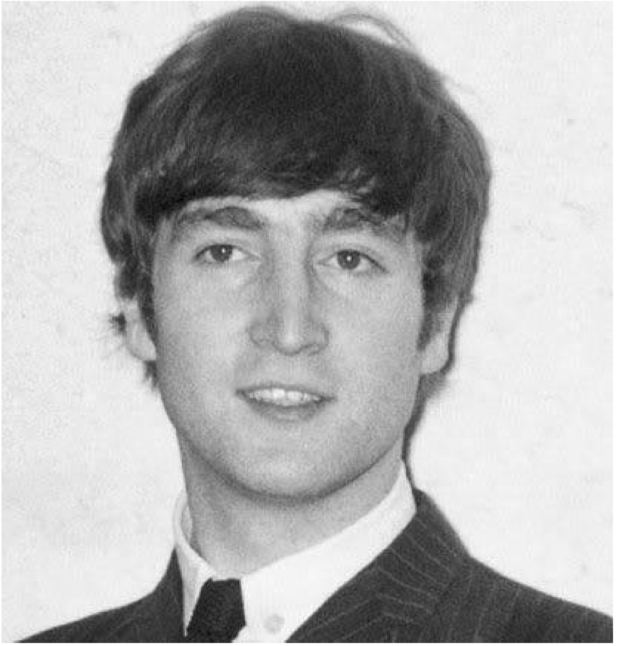
FRTE Misconception: Images are all high quality

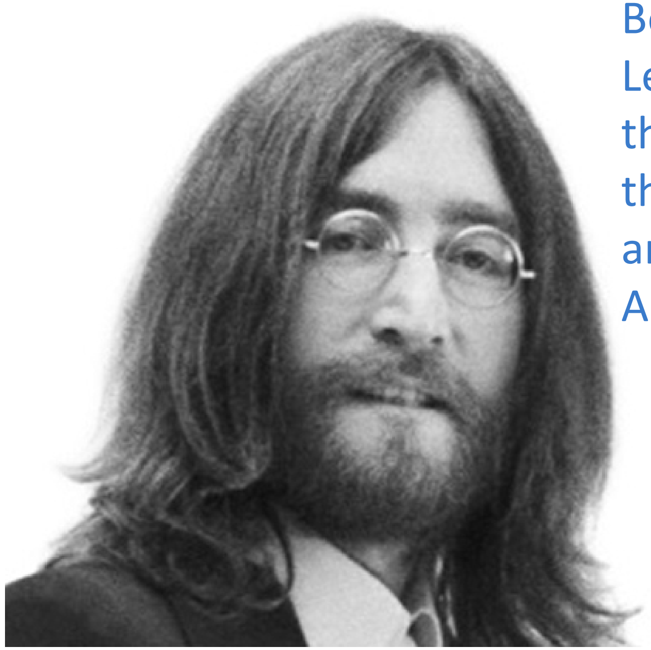
DHS southern border webcam photos.

Searched against mugshots in FRVT 1:N https://pages.nist.gov/frvt/html/frvt1N.html

AI Benefit :: Tolerance of appearance change





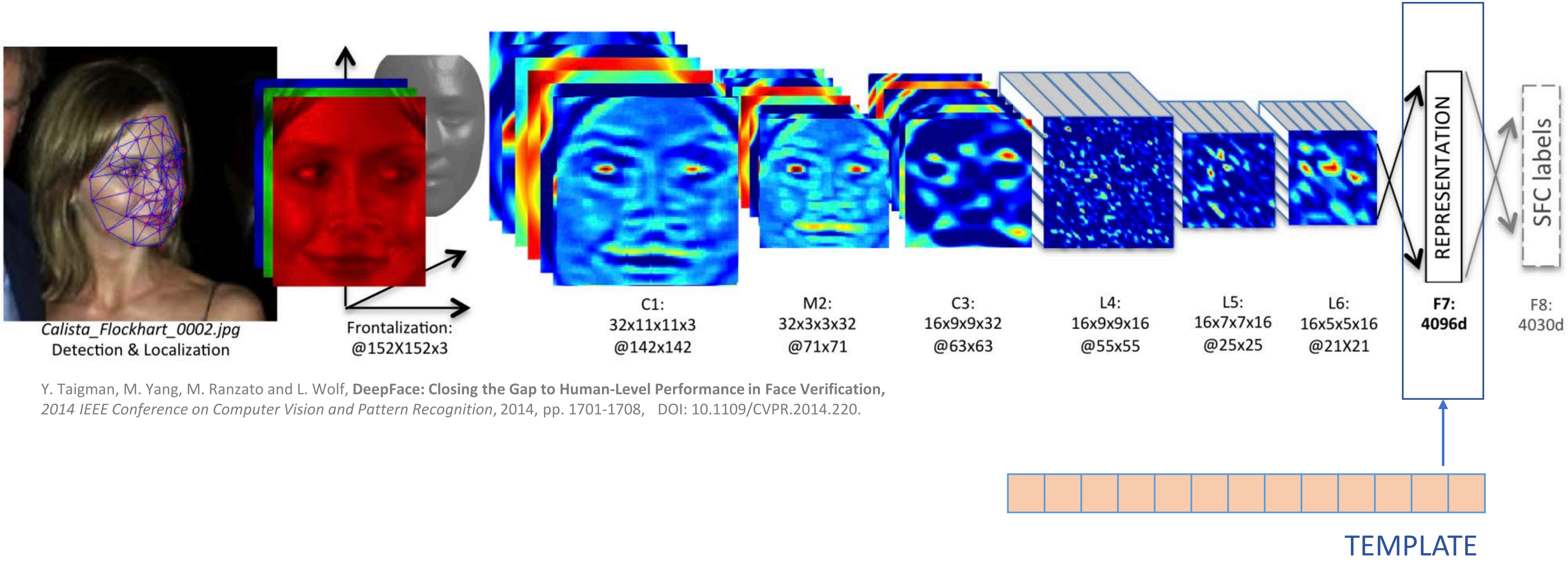


Beatle John Lennon between the release of the Red Album and the Blue Album, ~5 years.

Year	Developer	Algorithm	Score	FMR	Outcome
2021	Idemia	008	7438.78	< 5e-07	**Strong match**
2022	Paravision	010	0.38308	< 5e-07	**Strong match**
2014	Cogent Thales	A20A	2521	0.48	**Failed match**
2014	NEC	E20A	0.562	0.002	**Failed match**

FACEBOOK'S DEEPFACE (2014 = OLD!)

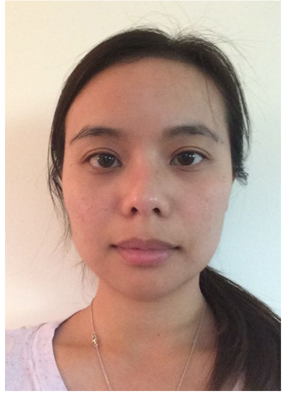
NIST



Y. Taigman, M. Yang, M. Ranzato and L. Wolf, **DeepFace: Closing the Gap to Human-Level Performance in Face Verification,** *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1701-1708, DOI: 10.1109/CVPR.2014.220.

FACE RECOGNITION MEASURES SIMILARITY OF FACES

NIST



1. Face Recognition (Identity Extractor)

1. Face Recognition (Identity Extractor)

2. Face Recognition Comparator

Template
aka Embedding
aka Feature vector

Similarity Score

Template
aka Embedding
aka Feature vector

No, different person! ← Is Score above threshold? → **Yes, same person!**

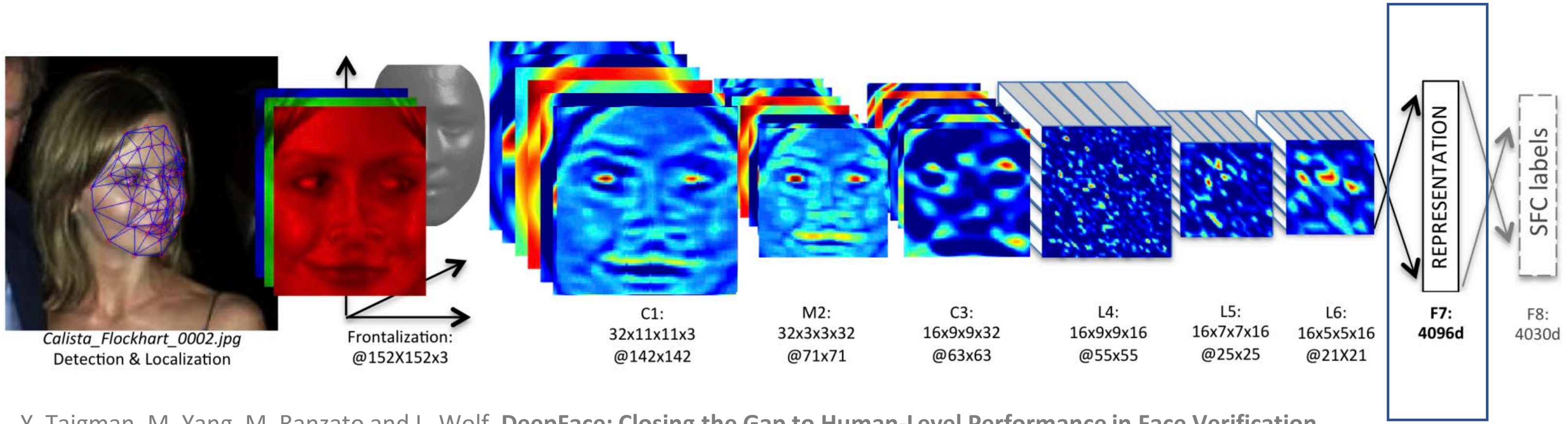
BOX 1: There's no standard for an FR Template

- Bespoke Neural Networks
- High-end ML / AI Intellectual property
- Trade secrets, black box
- Not commoditized

BOX 2:

- Not probabilities, not "percentage matches"
- Usually simple code, often fast
- No standards on output scores
- Various ranges [0,1] [0,100] [30-70] etc etc.

AI ⟶ Risks



Y. Taigman, M. Yang, M. Ranzato and L. Wolf, **DeepFace: Closing the Gap to Human-Level Performance in Face Verification,** *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1701-1708, DOI: 10.1109/CVPR.2014.220.

GENERIC AI TRUSTWORTHINESS

- Valid and reliable
- Fair
- Safe
- Secure
- Resilience
- Explainable, interpretable
- Privacy preserving

FACE TRUSTWORTHINESS

- Accuracy (FN and FP)
- Demographic effects small + manageable
- Cameras, environment
- Backdoors? Cybersecure?
- Correct function with anomalous inputs
- Rejection of attacks
- Courtroom testimony?
- Leakage? Cybersecurity? Hackable?

What constitutes “best” algorithm

- Accuracy
  - At small N vs. large N
  - Demographic dependence
- Time needed to make a template
- Time needed to search a database
  - At large N
  - Sublinear search
- Memory consumption
- Server | Embedded | Phone | Edge | Cloud
- SDK and API maturity, flexibility
- Forensic tools for investigation, clustering, GUI-based photo comparison
- Cost
  - Pricing model
  - Technology version refresh cost

Algorithm	Date	Memory (MB)	Template (B)	Template Time (ms)
cognitec-004	2022-02-10	585 (117)	2052 (331)	463 (115)
paravision-010	2022-02-02	2150 (374)	4100 (431)	634 (196)
rankone-013	2022-07-09	149 (27)	261 (6)	690 (226)
idemia-009	2022-07-27	2702 (396)	636 (61)	1207 (390)
cogent-007	2022-04-11	1884 (356)	550 (59)	1329 (423)
sensetime-007	2022-06-17	5699 (444)	1028 (79)	1386 (435)

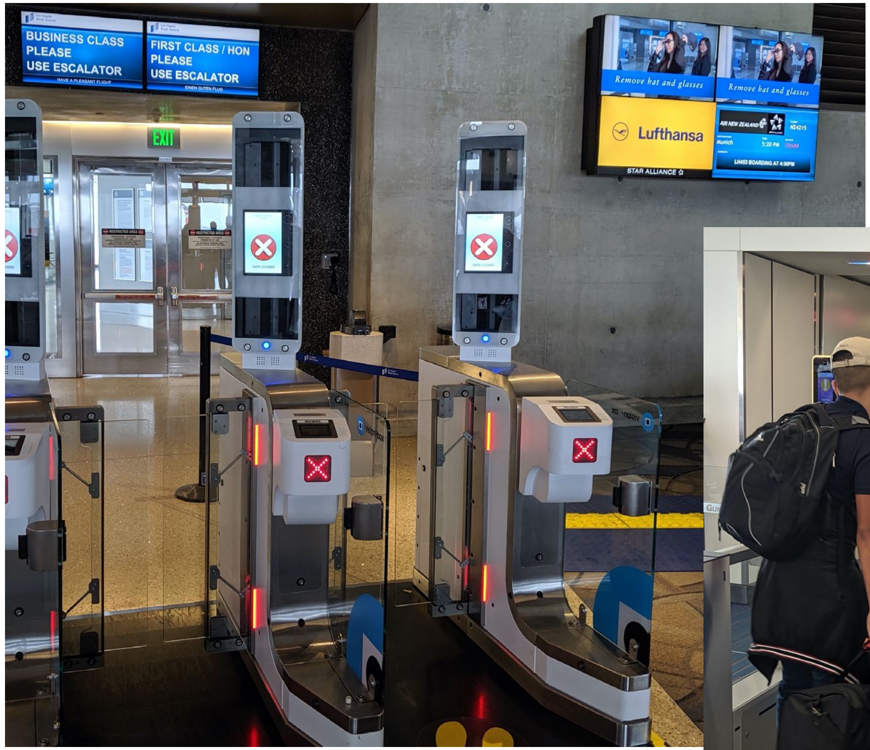
Memory: 148MB vs. 5.6GB

Template time: 0.5 secs vs. 1.4 secs

Failed Capture / Quality Assessment / Downstream Consequences

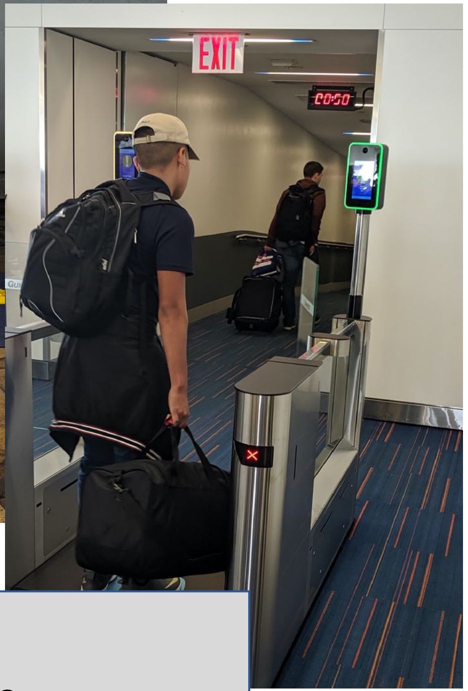
Source: https://www.cbp.gov/newsroom/local-media-release/cbp-introduces-simplified-arrival-denver-international-airport

Face detector? No
Face quality assessment? No
Failure-to-capture rate = 0
⟶ but FNMR greater downstream



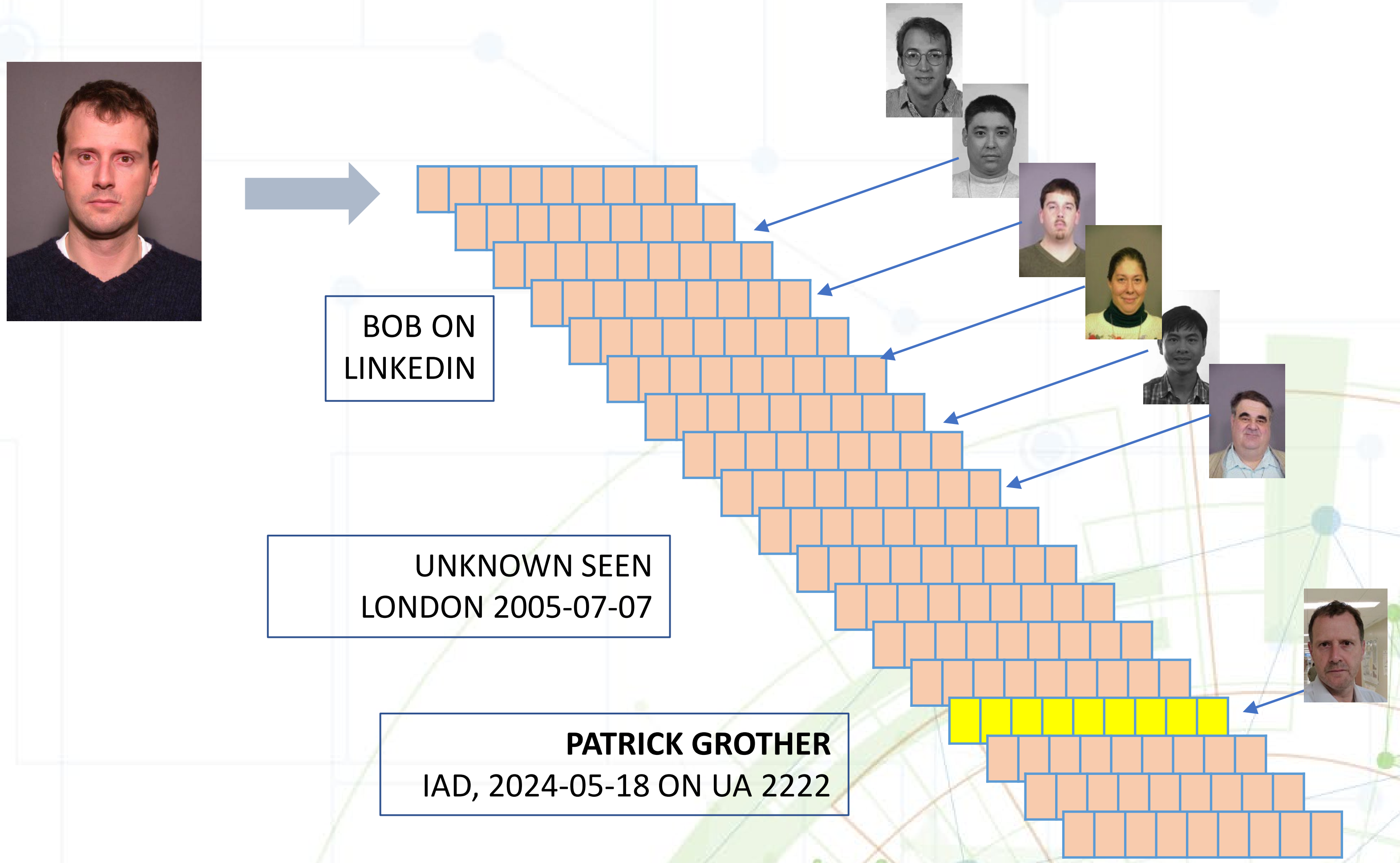
Source: NIST Staff at LAX

Source: NIST Staff at MCO



Face detector? Yes
Face quality assessment? Yes
Failure-to-capture rate > 0
⟶ so FNMR reduced downstream

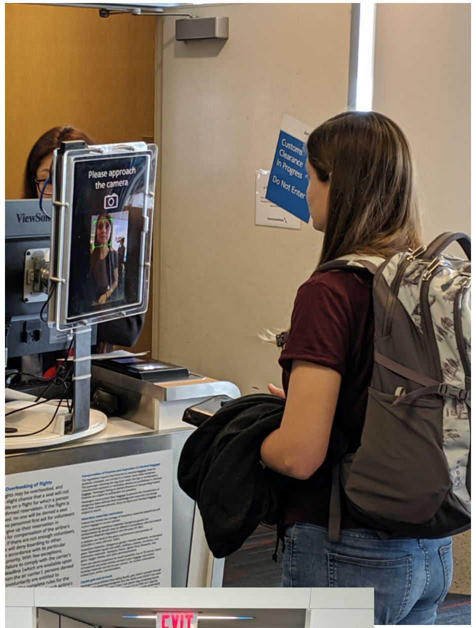
THE STATE OF THE ART 1:N SEARCH

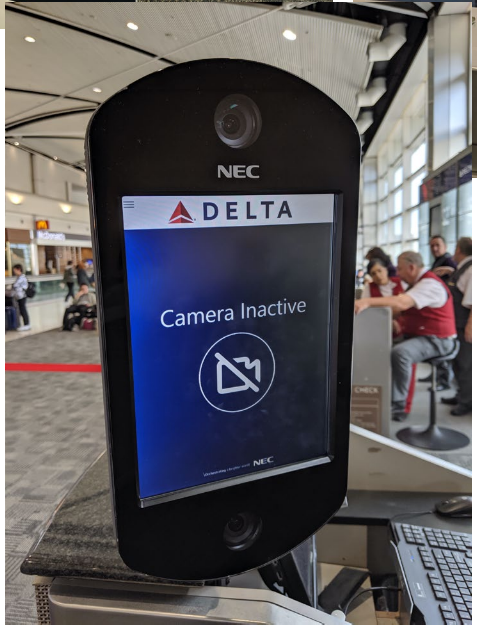


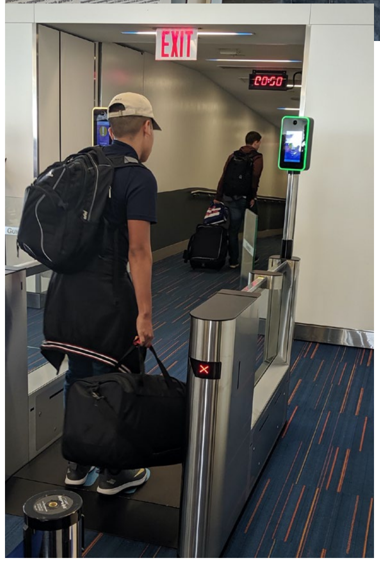
POSITIVE IDENTIFICATION IN ACTION :: DEPARTURES FROM USA

DOUBLE DUTY:
1. POSITIVE ACCESS CONTROL
2. IMMIGRATION EXIT FACILITATION

HOW:
1. FACE RECOGNITION 1:N
2. PAPERLESS BOARDING



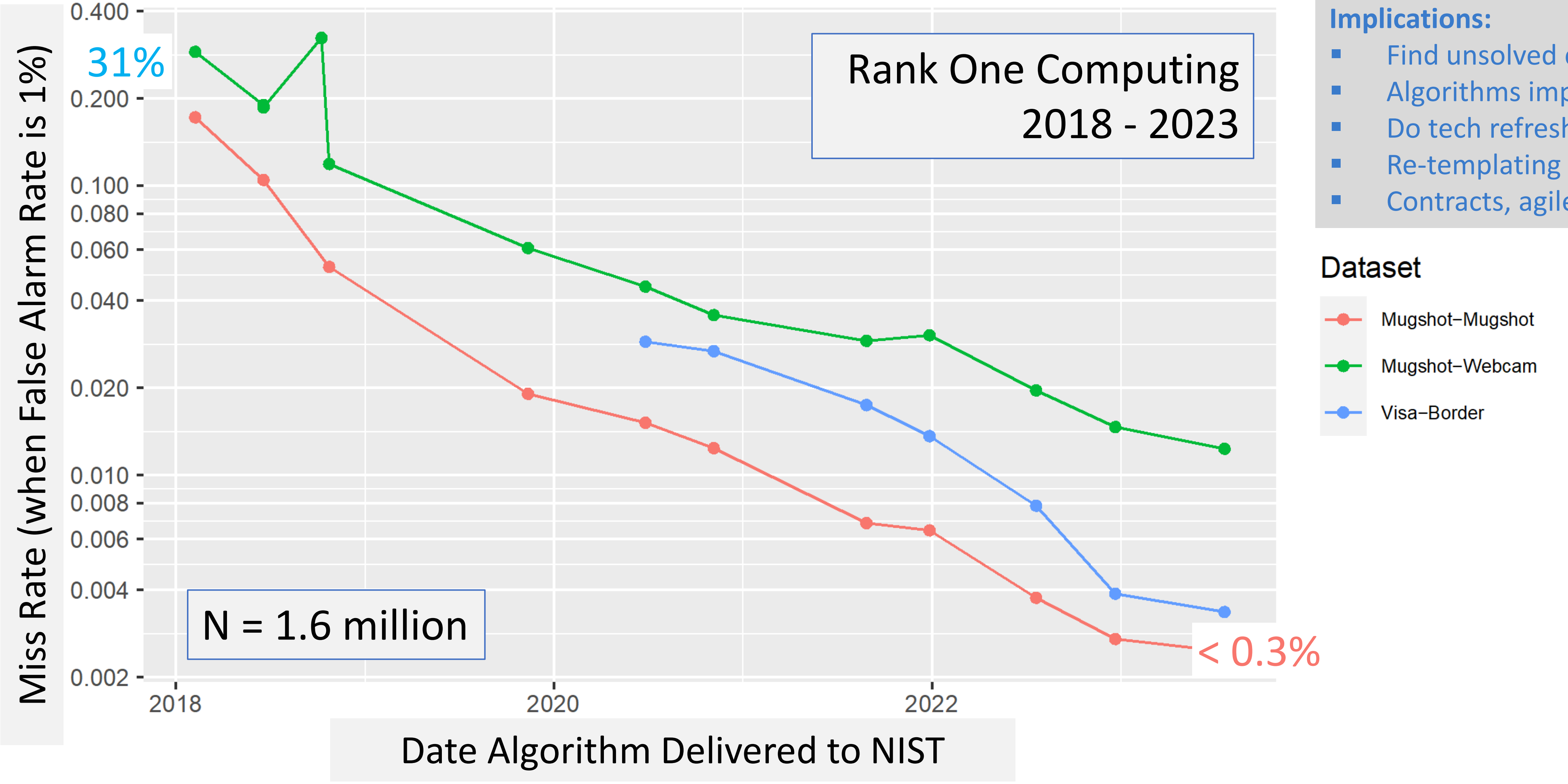




Diverse hardware, common matcher (TVS)

1:N ACCURACY GAINS CONTINUE

NIST



Implications:

- Find unsolved case leads
- Algorithms improve regularly
- Do tech refresh!
- Re-templating necessary
- Contracts, agile procurement

The Power of 1:N AFR Today

Enroll portrait into gallery with N = 12 million other people

SEARCH PHOTOS GIVING HIGH-SCORE MATE AT RANK 1

2002

2018

2007
50º YAW

2019

2008
SUNGLASSES

2014
60º YAW

2009
SHADOW

The Power of 1:N AFR Today

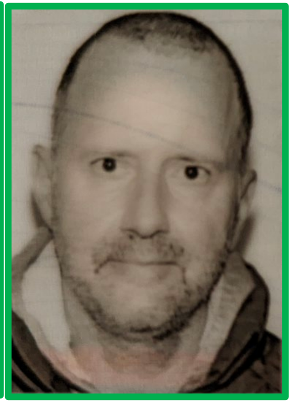
Enroll portrait into gallery with N = 12 million other people

SEARCH PHOTOS GIVING HIGH-SCORE MATE AT RANK 1

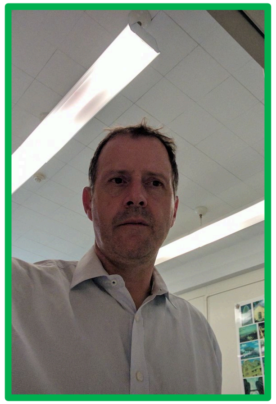
2003
SHADOW

2009
POSE

2007
55º PITCH



2021
DL

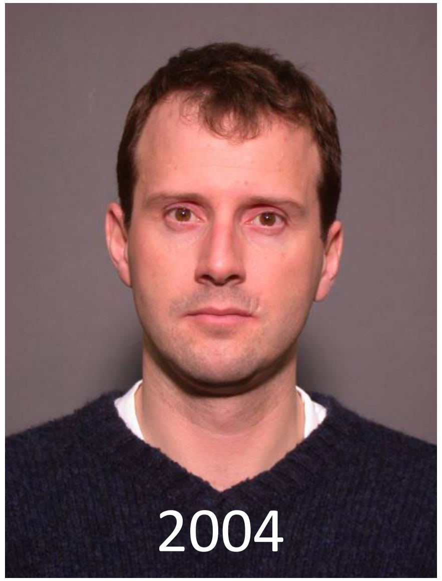


2019
PITCH

2014
90º YAW

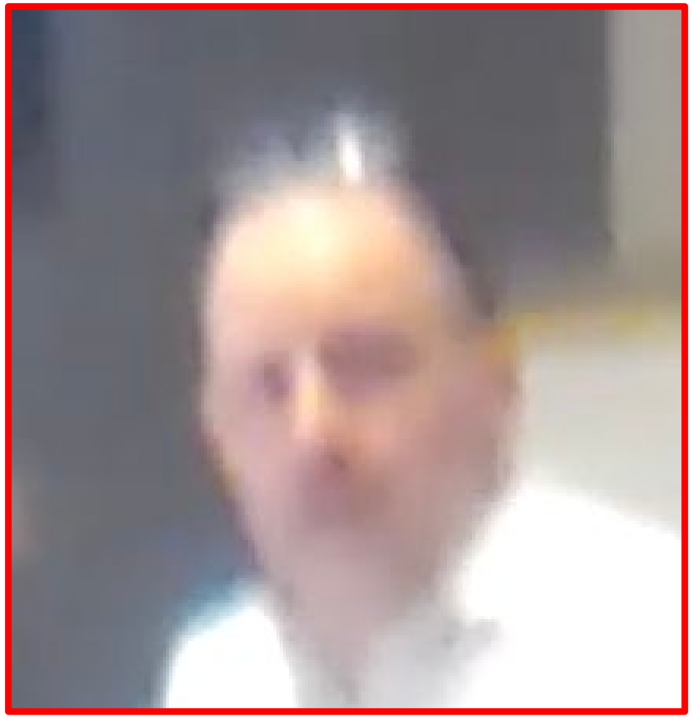
2014
40º PITCH

The Power of 1:N AFR Today



Enroll portrait into a gallery with N = 12 million other people

RANK 1 WEAK HIT



MISSED OUTSIDE 50 RANKS

HIT BY ONE CHINESE ALGORITHM RANK 15

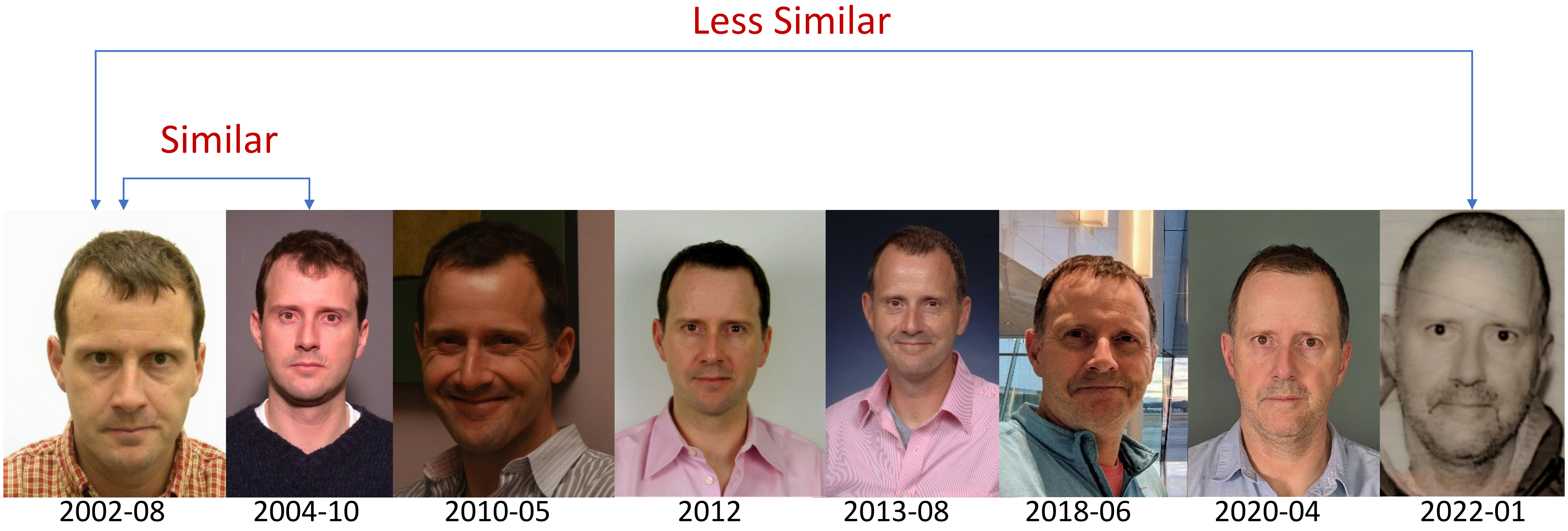
1:N False Positives

MEI
IN GALLERY WITH N = 12
MILLION MUGSHOTS

MEI'S SISTER

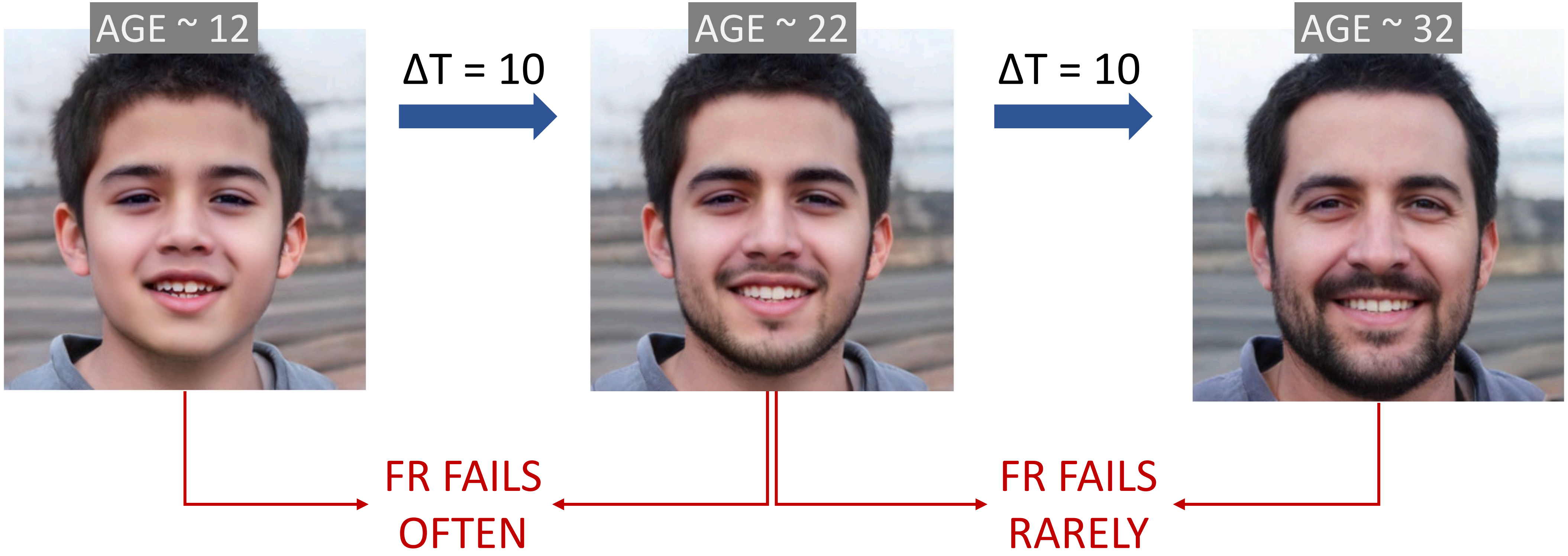
- 10 ALGS FIND GALLERY MATCH AT RANK 1, WEAK SCORE

AGEING: APPEARANCE CHANGES ⟶ REDUCED SIMILARITY



Images from presenter

CHILDREN HAVE RAPID CHANGE IN APPEARANCE



Images from: https://www.nytimes.com/interactive/2020/11/21/science/artificial-intelligence-fake-people-faces.html

AGEING: SEARCH ERROR RATES

AGE AT ENROLLMENT / PROBE TAKEN 10 YRS LATER

AGE AT ENROLLMENT	FEMALE	MALE
(60:99]	0.019	0.009
(45:60]	0.018	0.005
(30:45]	0.021	0.006
(21:30]	0.030	0.011
(18:21]	0.064	0.037
(15:18]	0.105	0.119
(12:15]	0.155	0.223

TAKEAWAYS:

1. 10 YEARS IS A LONG TIME FOR A TEEN
2. MOST ACCURATE AGES 30 TO 60
3. MEN EASIER TO RECOGNIZE
   - EXCEPT IN TEEN YEARS
4. SOME ALGORITHMS MUCH BETTER

MISS RATES:
15.5% FEMALE VS. 22.3% MALE

Algorithm: Clearview AI 2024-02

Gallery: AIR-ENTRY, N = 1.6 million, balanced by sex and by specific age-groups

Probes: AIR-ENTRY, 3.8 million searches, balanced by sex, age-group and time-lapse (years)

COMPARATIVE ACCURACY

NIST

AGE AT ENROLLMENT

NEC-2023-12

	FEMALE	MALE
(60:99]	0.006	0.003
(45:60]	0.003	0.002
(30:45]	0.004	0.002
(21:30]	0.004	0.003
(18:21]	0.006	0.003
(15:18]	0.009	0.009
(12:15]	0.014	0.014

PROBE TAKEN 10 YRS LATER

IDEMIA-2024-03

	FEMALE	MALE
(60:99]	0.018	0.006
(45:60]	0.017	0.004
(30:45]	0.019	0.005
(21:30]	0.027	0.008
(18:21]	0.052	0.024
(15:18]	0.082	0.076
(12:15]	0.111	0.156

PROBE TAKEN 10 YRS LATER

CLEARVIEW AI -2024-02

	FEMALE	MALE
(60:99]	0.019	0.009
(45:60]	0.018	0.005
(30:45]	0.021	0.006
(21:30]	0.030	0.011
(18:21]	0.064	0.037
(15:18]	0.105	0.119
(12:15]	0.155	0.223

PROBE TAKEN 10 YRS LATER

ONE-TO-MANY MISS RATES BY AGE AND AGEING

Slow increase between 10-15 years

Higher error rates in teenagers: Rapid ageing

AGE AT ENROLLMENT

TIME LAPSE BETWEEN SEARCH AND ENROLLMENT

Female

	10	11	12	13	14	15
(60:99]	0.014	0.017	0.022	0.022	0.026	0.024
(45:60]	0.013	0.013	0.015	0.019	0.022	0.021
(30:45]	0.014	0.015	0.018	0.022	0.025	0.025
(21:30]	0.018	0.019	0.022	0.027	0.029	0.029
(18:21]	0.034	0.034	0.038	0.046	0.044	0.049
(15:18]	0.054	0.049	0.057	0.064	0.070	0.066
(12:15]	0.080	0.070	0.076	0.089	0.106	0.109

Male

	10	11	12	13	14	15
(60:99]	0.008	0.008	0.009	0.012	0.009	0.010
(45:60]	0.006	0.007	0.008	0.010	0.011	0.012
(30:45]	0.007	0.006	0.008	0.010	0.012	0.013
(21:30]	0.010	0.010	0.011	0.014	0.017	0.019
(18:21]	0.030	0.031	0.036	0.041	0.053	0.046
(15:18]	0.090	0.091	0.101	0.132	0.145	0.150
(12:15]	0.174	0.160	0.206	0.237	0.252	0.246

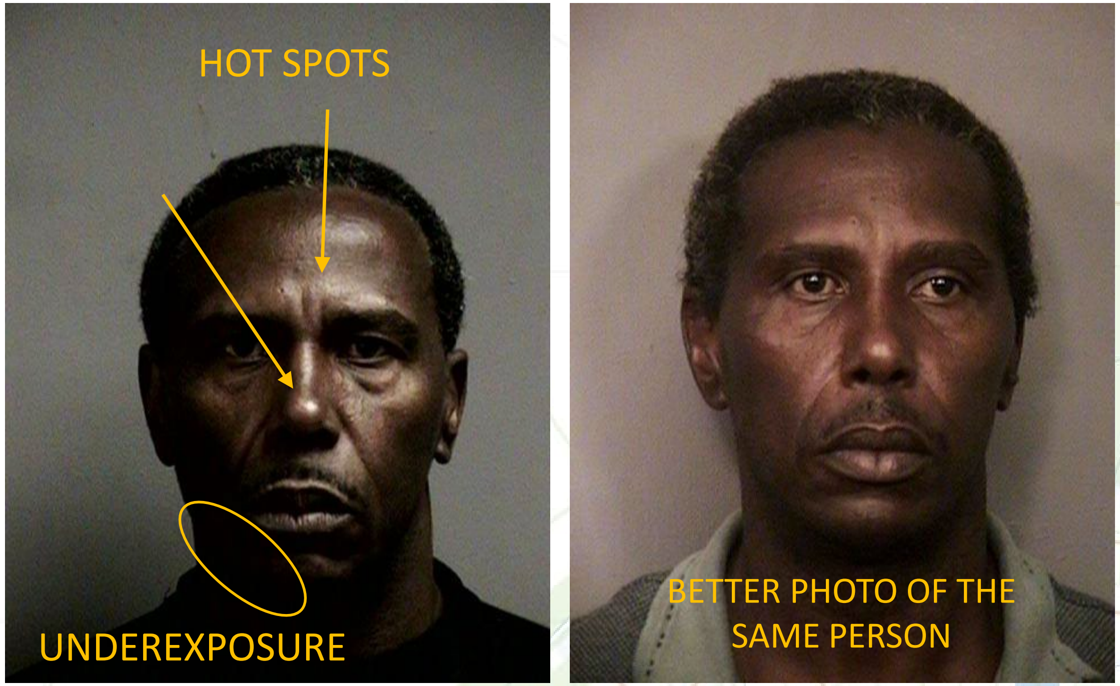
Algorithm: Panasonic 2023_08

Gallery: AIR-ENTRY, N = 1.6 million, balanced by sex and by specific age-groups

Probes: AIR-ENTRY, 3.8 million searches, balanced by sex, age-group and time-lapse (years)

DEMOGRAPHICS #1

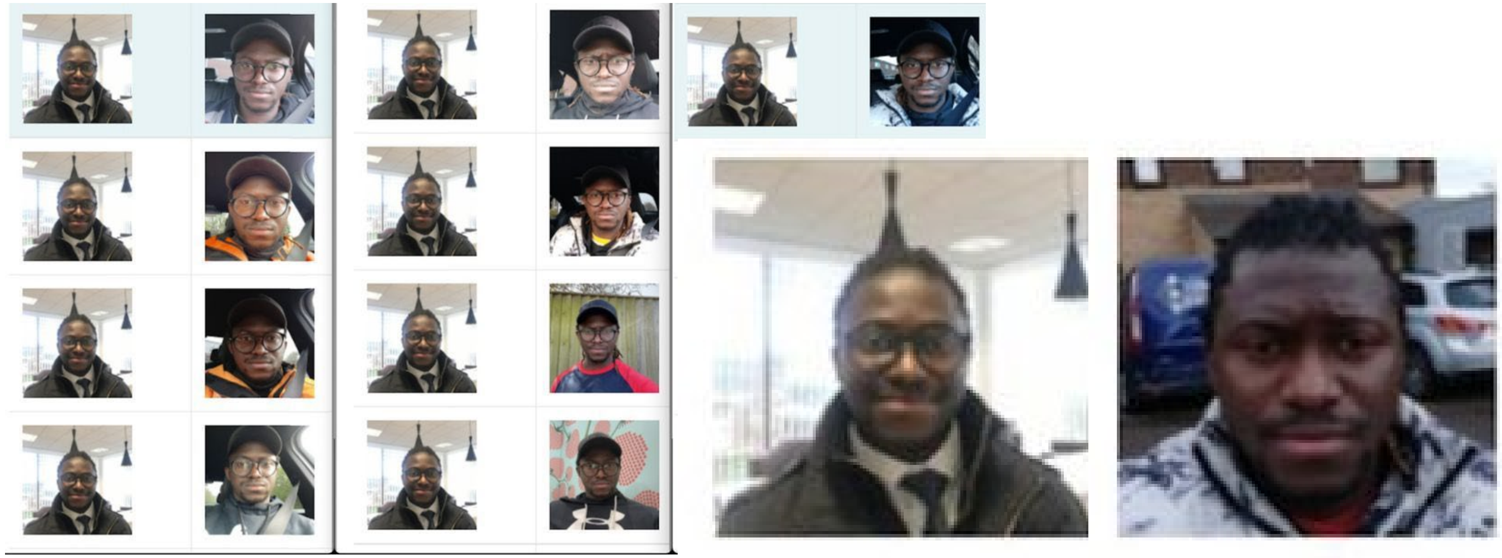
Do some groups have higher failure-to-match rates?



Source: NIST Special Database 32 aka “MEDS”, subject S171

Demographics: A False Negative Anecdote

NIST



Source: https://www.adcu.org.uk/news-posts/uber-facial-recognition-discrimination

Respondent(s) of Microsoft facial recognition software. This requires drivers to take a real time photograph of themselves (a 'selfie') for verification when using the app. The photograph is then checked against the driver's account profile picture.

Pleadings of Pa Edrissa Manjang linked from https://www.adcu.org.uk/news-posts/uber-facial-recognition-discrimination

"The system includes robust human review to make sure that we're not making decisions about someone's livelihood in a vacuum, without oversight," the [Uber] spokesperson said.

https://www.uktech.news/ai/uber-eats-racist-ai-dismissal-20220728

Couriers say Uber's 'racist' facial identification tech got them fired

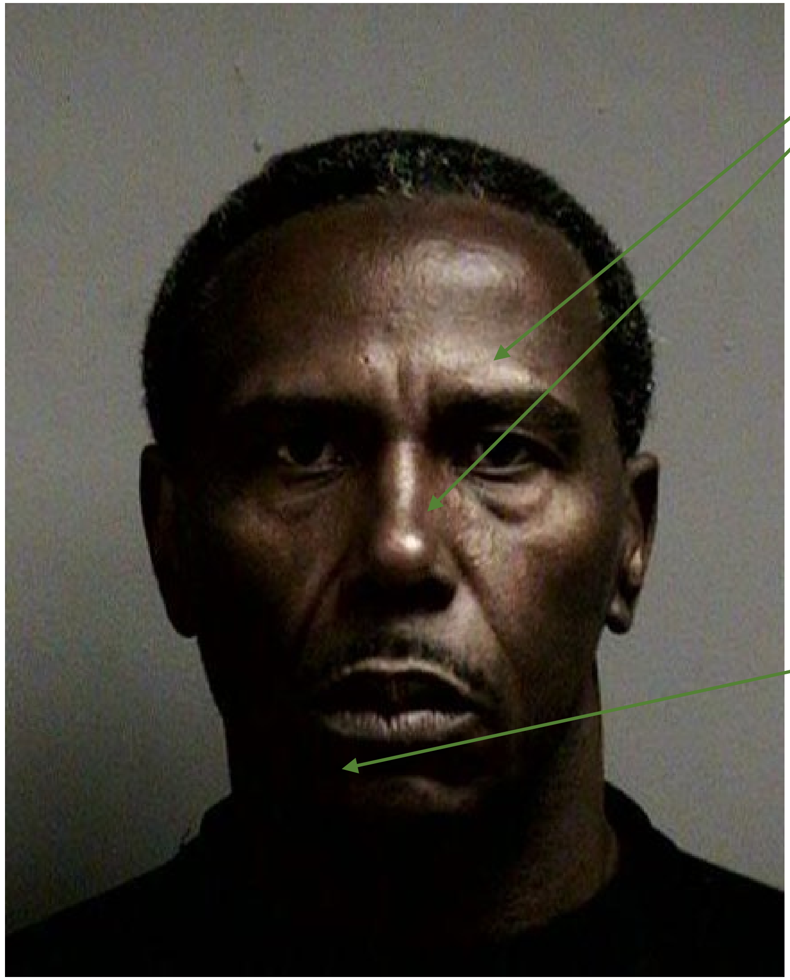
BAME couriers working for Uber Eats and Uber claim that the company's flawed identification technology is costing them their livelihoods

GETTY IMAGES / WIRED

Uber Eats couriers say they have been fired because the company's "racist" facial identification software is incapable of recognising their faces. The system, which Uber describes as a "photo comparison" tool, prompts couriers and drivers to take a photograph of themselves and compares it to a photograph in the company's database.

https://www.wired.co.uk/article/uber-eats-couriers-facial-recognition

One source of false negative bias: Photography



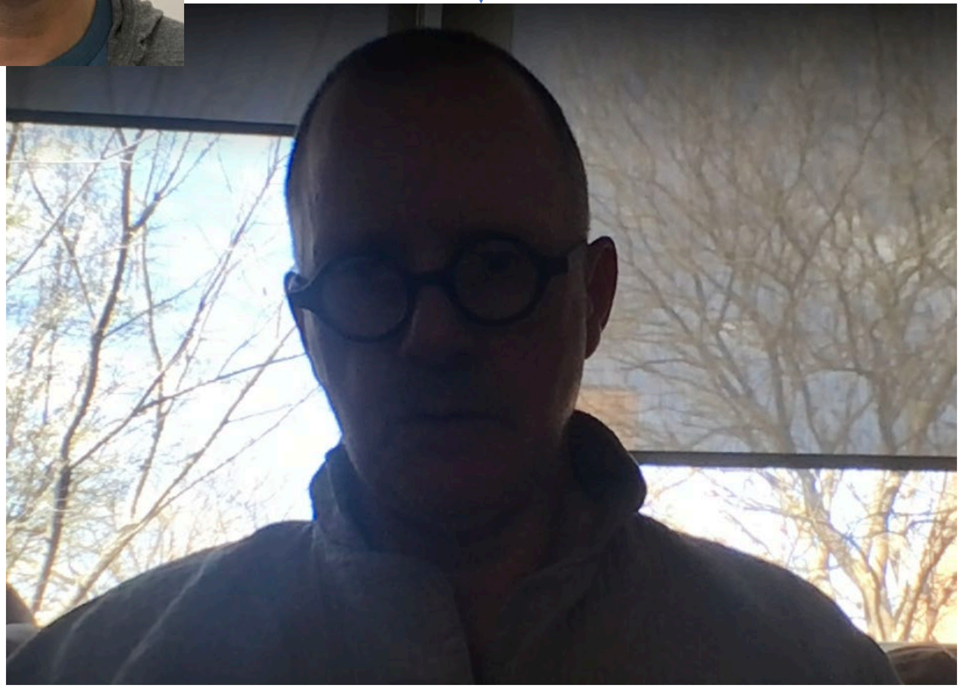
Example of an underexposed photo
from NIST Special Database 32

A. The photograph contains **specularities**, bright areas due to the surface orientation of the skin.
B. Dark skin reflects less light so there is high contrast between the specular and diffuse reflection areas.
C. Light skin reflects more light so across the face **contrast is relatively low**.
D. Many cameras convert incident light into digital images with a 256 level data type that does not allow the full range of reflected light to be represented.
E. This can result in **underexposure** of subjects with dark skin where information used by recognition algorithm is reduced or absent.
F. Some face-aware cameras can use high dynamic range imaging, computational photography, and AI to ameliorate this problem.

Example of an overexposed photo
from NIST Special Database 32

POOR PHOTOS: CAMERA - ENVIRONMENT INTERACTION

FR comparison fails: False Negative

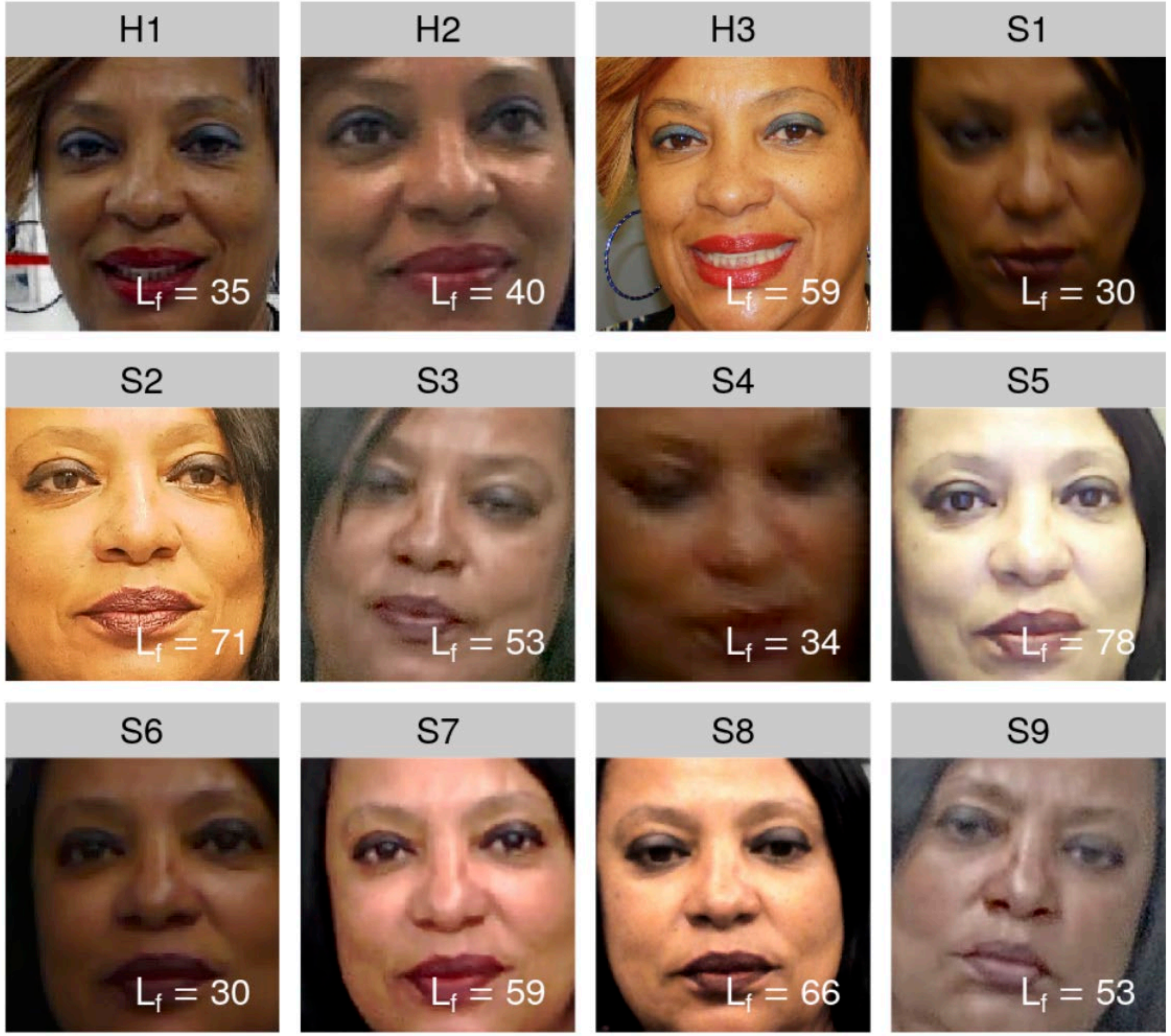


Underexposure

FR comparison succeeds: True Positive

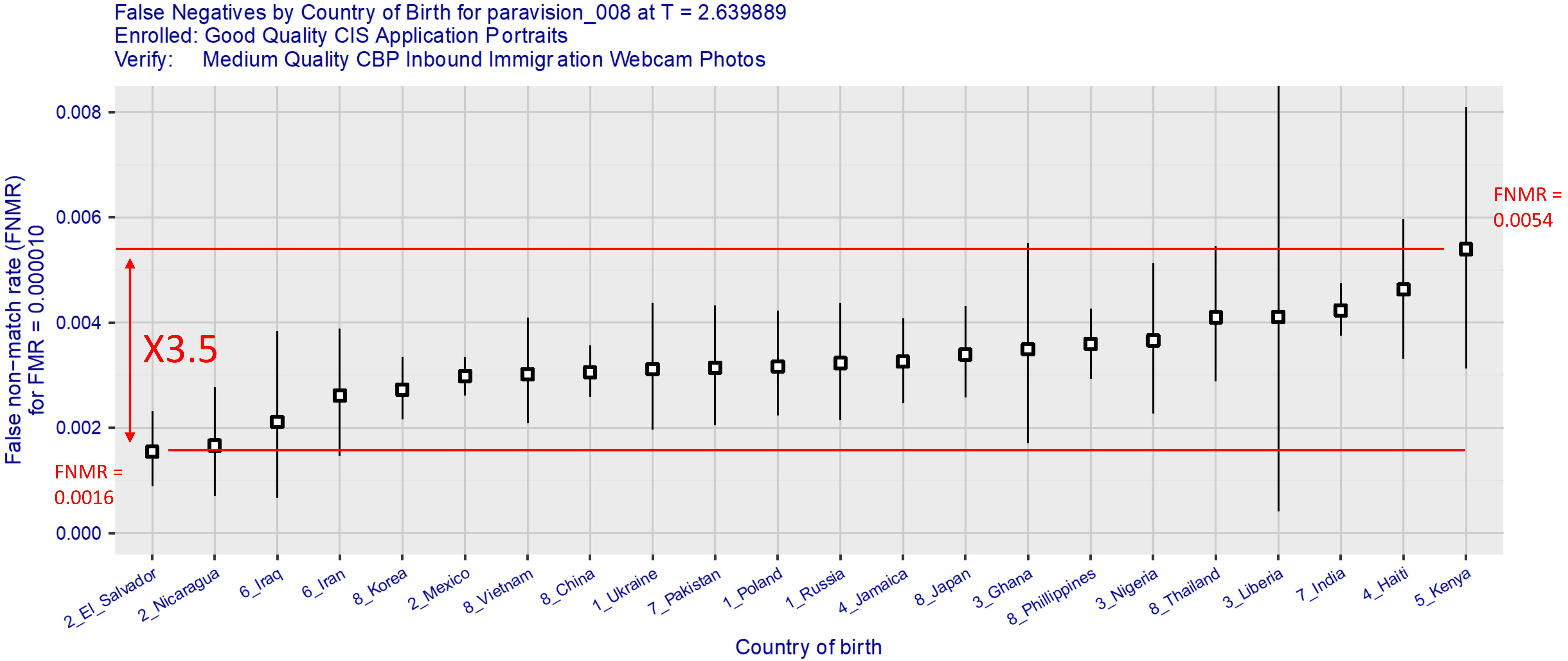
Better exposure

Two people, each imaged by multiple cameras, minutes apart



Reliability and Validity of Image-Based and Self-Reported Skin Phenotype Metrics
John J. Howard, Yevgeniy B. Sirotin, Jerry L. Tipton, and Arun R. Vemury
https://mdtf.org/publications/arXiv2021-FALMs.pdf, 2021 (and IEEE BIOM, to appear)

False Negatives by Country of Birth (Competitive US Algorithm)



- FNMR is absolutely small – better than 0.5%
- FNMR ratio across “racial groups” is small around 3 times

Impact of false negative dependence on demographics

1 How false negatives occur

False negatives from low mate scores from change in appearance

- Image defects aka quality
- Ageing
- Injury
- Cosmetics

2 When does this occur

When most of your transactions are mated

- Access control
- Time and attendance
- Border crossing
- Drug dispensing

When you don't own, or have control of, the capture process

3 Poor photography

Dark skin-tone can make photography difficult

Not the algorithm's fault

- Bad photography = Garbage In, Garbage Out
- But algorithm developers need to understand their neural network's response to poor exposure

4 Who?

In populations with mixed demographics

- Children
- Women
- Dark skin, fair skin
- Tall people, short people
- Wheelchair-bound people

FACE-AWARE CAPTURE

Google’s Real Tone (in Pixel 6+ phones)

PIXEL

Image equity: Making image tools more fair for everyone

Oct 19, 2021
3 min read

As part of Google’s Product Inclusion efforts, our teams are building more equitable camera and imaging products for people of color.

Florian Koenigsberger
Google Image Equity Lead

Share

Pictures are a big part of how we see each other and the world around us, and historically racial bias in camera technology has overlooked and excluded people of color. That same bias can carry through in our modern imaging tools if they aren’t tested with a diverse

- **https://store.google.com/intl/en/ideas/real-tone/**
- https://blog.google/products/pixel/image-equity-real-tone-pixel-6-photos/
- https://blog.google/inside-google/company-announcements/super-bowl-ad-2022/

DEMOGRAPHICS #2
Do some groups have higher mis-match rates?

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

Apple’s Note on False Match Rates

Apple Face ID claims **FMR ~ 1:1 000 000**

https://support.apple.com/en-us/HT208108
Retrieved 2024-04-22

“The statistical **probability is higher**—and further increased if using Face ID with a mask—**for twins and siblings** that look like you, and among children under the age of 13, because their distinct facial features might not have fully developed.”

From Apple’s iPhoneX demo September 12, 2017

UNDER 13

https://freephotos.cc/

SIBLINGS

NIST staff + sister, with permission

THE MUCH BIGGER PROBLEM: FALSE POSITIVE RATE VARIATION

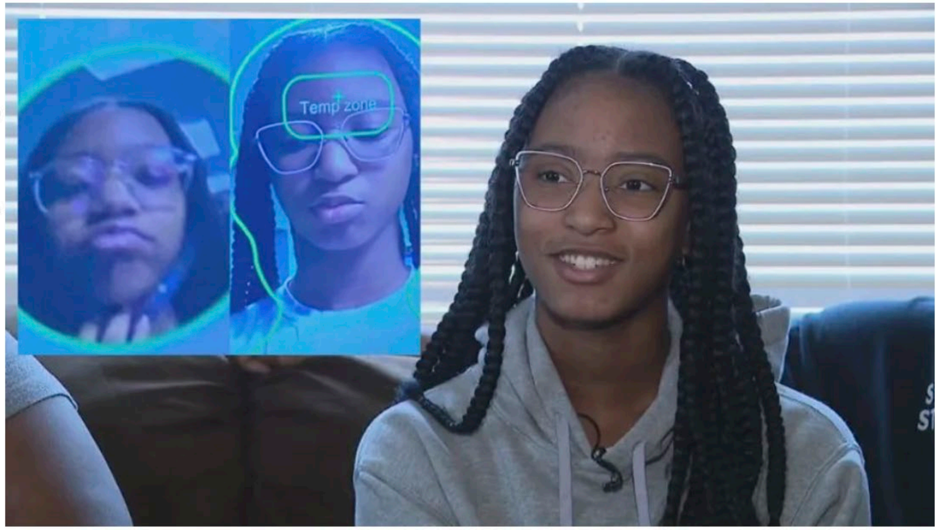
NIST

A: Many more false positives in
- Women
- Ethnicities unknown to the algorithm
- The very young, and old

B: Critical in search applications like
- Casinos
- Football stadiums
- Big brother surveillance
- Duplicate detection

BLACK GIRL BANNED FROM MICHIGAN SKATING RINK BECAUSE FACIAL RECOGNITION SOFTWARE MISIDENTIFIED HER

by Cedric 'BIG CED' Thornton July 16, 2021 4948



(Image: Fox 2 Detroit)

A young Black girl was kicked out of and banned from a skating rink in Michigan through no fault of her own. The girl was been banned due to facial recognition software that misidentified her as someone else.

https://www.zdnet.com/article/backlash-to-retail-use-of-facial-recognition-grows-after-michigan-teen-kicked-out-of-skating-rink-after-false-match/

Demographics Summary

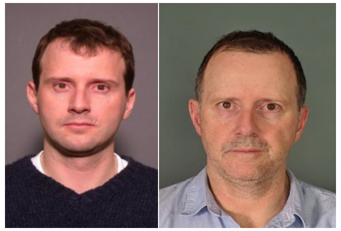
- **Leading algorithms today**
  - Are very accurate
  - Increasingly tolerate poor image quality
  - But errors unequal across demographics
- **Tests show**
  - **False positive differentials >> false negative differentials**
  - More false positives in Asian and African faces
  - More false positives in women
  - More false positives in the old and very young
- **One-to-many algorithms** do not behave like one-to-one
  - Many do
  - But some one-to-many stabilize false alarm rates

❑ **False negatives from bad photography**

❑ **False positives from algorithms applied to “unknown” demographic groups (even with high quality images)**

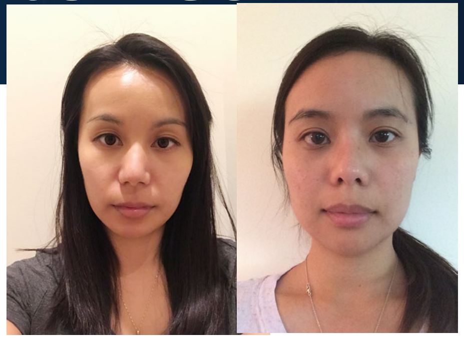
- **Know-Your-Algorithm, Know-Your-System**
  - Accuracy
  - Demographic sensitivity
  - Threshold to limit false positives on worst-case demographic
  - Traceability to (NIST) tests
- **So what? It depends on the application**
  - Error impacts range from grave to inconsequential.
- **Incomplete reporting** in the press
  - Confusion of face “analysis” with “recognition”
  - Don’t say which component is at fault
  - Don’t differentiate false positives from false negatives
  - Missing reports on false positives
- **Gains**
  - Some developers have attempted to address differentials.
  - We have summary indicator
  - Academic research

Demographics in AFR: Two separate stories



FALSE NEGATIVES DIFFERENTIALS

- FN involves two images of one person
- FN occurs when the similarity score is low
- Low similarity occurs if change in appearance between the images
  - Examples of change (left): Ageing, hairstyle, reduced info as dark skin reflects less light (physics), cosmetics.
- Empirical results:
  - Higher FN in women, Africans and African Americans.
  - Effects are variable across algorithms. Most accurate algorithms, generally give lowest differences in FNMR.
  - FNMR is generally low, with factor of 3 span across demographic groups.
- Responsible party for fixing this: Photographers, capture system and camera providers, quality algorithm developers
  - Also use more capable algorithm
- Worst affected applications: Applications where photography cannot be controlled, rapid capture in adverse enviros. Access control, benefits authentication
- Impact: Inconvenience or worse for one person



FALSE POSITIVE DIFFERENTIALS

- FP from one image from each of two people
- FP occurs when the similarity score is high
- High similarity score when appearance is similar (see above)
- Empirical results:
  - High FMR in women, the elderly, E. Asians, Africans, and S. Asians, and highest in, for example, elderly asian females.
  - But some Chinese-developed algorithms lowest FMR on E. Asians
  - Even with pristine well photographed images
  - FMR can span three orders of magnitude (x1000)
- Responsible party for fixing this: Algorithm developers
- How: More diverse training data, loss functions that force evenly clustered but separated demographic groups
- Worst affected applications: High volume applications with big galleries where most searches are non-mated e.g. watchlists, duplicate detection searches
- Impact: A FP can adversely affect either or both people

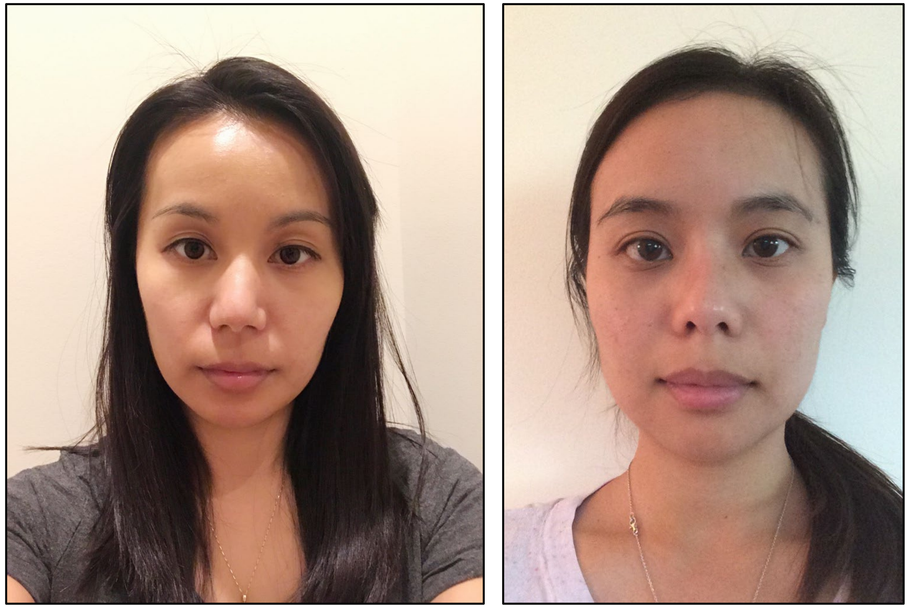
TWINS :: THE FALSE POSITIVE PROBLEM

SOURCE: TWINS DAY OHIO COLLECTED BY NOTRE DAME

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

AFR MIS-MATCHES ON TWINS (AND SIBLINGS)

Source: Notre Dame's Twins Day Collection



Source: Mei Ngan and her sister

Developer	Algorithm	Score	FMR	Outcome
IDEMIA	009	4924.38	< 5.049e-07	FALSE MATCH!
PARAVISION	010	0.322402	< 5.049e-07	FALSE MATCH!

SAME PERSON OR NOT?

Source: Notre Dame’s Twins Day Collection

	Identical	Fraternal
How	Monozygotic	Dizygotic
USA proportion that are a twin	0.7%	3.1%
West Africa	0.5%	2.8%
East Asia	0.3%	0.9%
Same-sex	100%	50% in theory 58% actually
Twinning rate	x1.5 since 1980	x1.9 since 1980
Demographics	~ constant with age, geography	varies with mothers age, order, geography

Twins, triplets ... constituted 123,000 out of 3.7M births in US 2019 - https://www.cdc.gov/nchs/data/nvsr/nvsr70/nvsr70-02-508.pdf

HUMAN CAPABILITY

CHOICE A: Same person

CHOICE B: Different person

- HUMAN COMPARISON OF PHOTOS IS HARD!
- TRY IT YOURSELF via UNI. NEW SOUTH WALES

FAMILIAR FACES

Source: Pixabay.com

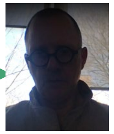
SLIDE FROM Frøy Løvåsdal
Team Identity, biometrics and biometric data EUIS-programme
National Police Directorate, Norway
froy.lovasdal@politiet.no

POLITIET
POLITIDIREKTORATET

UNFAMILIAR FACES

SLIDE FROM Frøy Løvåsdal
Team Identity, biometrics and biometric data EUIS-programme
National Police Directorate, Norway
froy.lovasdal@politiet.no

INVESTIGATION VS. (LIGHTS OUT) IDENTIFICATION



Search image

Score	Rank
3.142	1
2.998	2
1.626	3
0.707	4
0.330	5
0.198	6
0.074	7
0.016	8

MATE

Policy A: Review up to 8 candidates

Source: https://www.google.com/doodles/jim-hensons-75th-birthday

"Investigation"
Returns candidates for human review

"Lights Out"
System concludes this the correct result

Search image

Score	Rank
3.142	1

T

Policy B: Review only candidates with score above threshold T = 3.0

Incorrect Arrests in Michigan and New Jersey

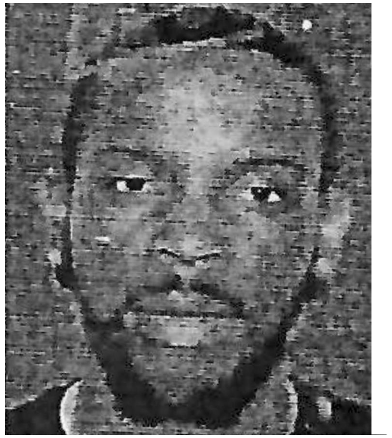
PROBE

INCORRECT PERSON
ROBERT WILLIAMS

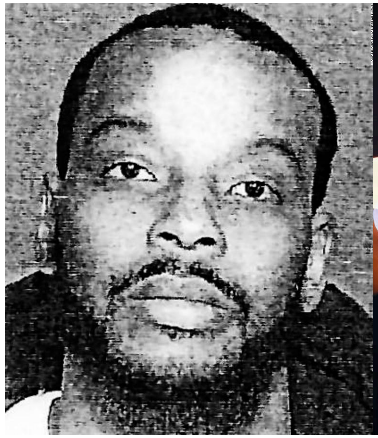
https://www.cbsnews.com/news/facial-recognition-60-minutes-2021-05-16/

INCORRECT PERSON
NIJEER PARKS

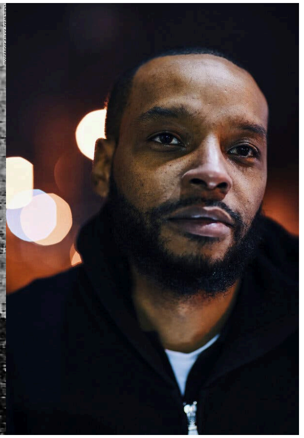
PROBE



GALLERY RETRIEVED



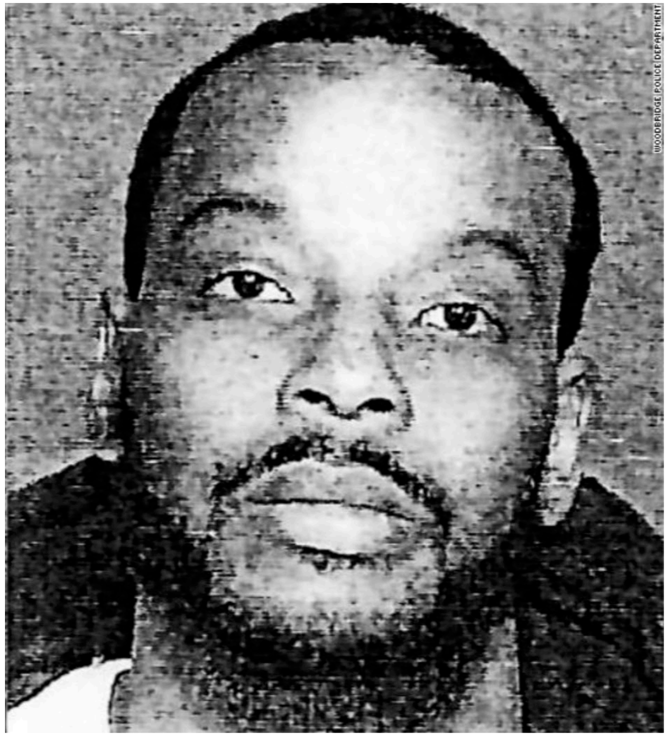
NEWS PHOTO



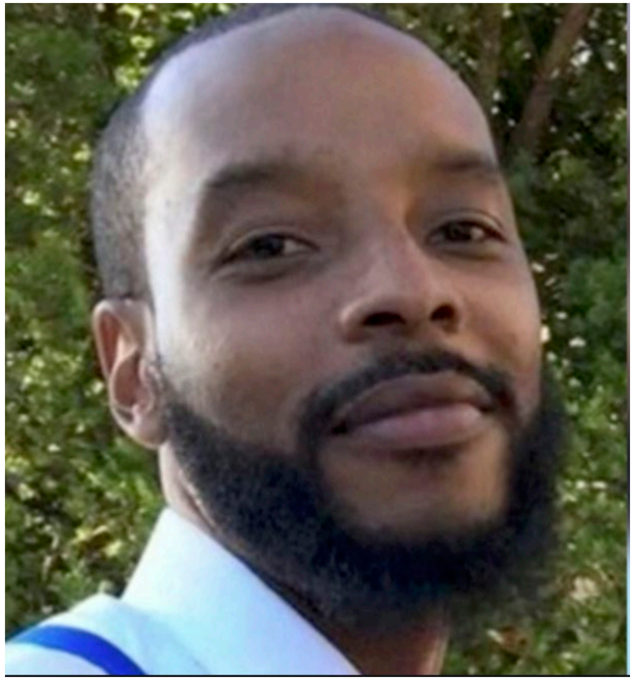
https://www.cnn.com/2021/04/29/tech/nijeer-parks-facial-recognition-police-arrest/index.html

https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html

FR RETURNS CLOSE BUT INCORRECT CANDIDATES

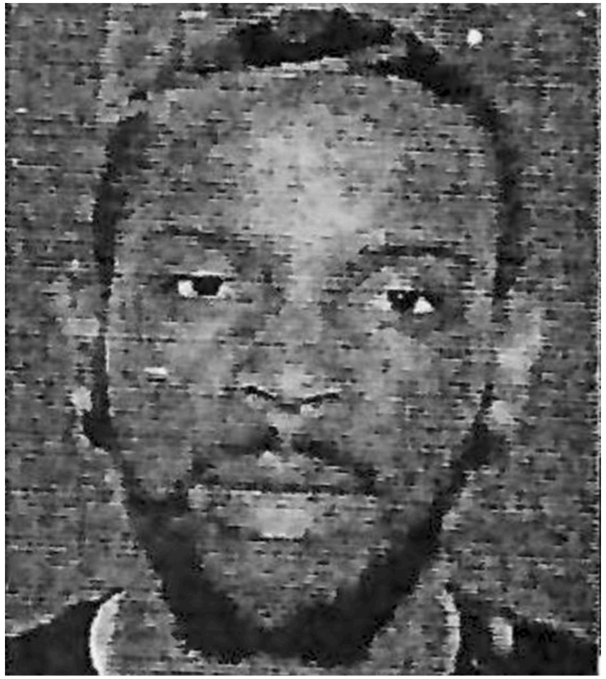


NIJEER PARKS
IN GALLERY, N = 12M.



NIJEER PARKS
POST EXONERATION

- 10 ALGS FIND GALLERY MATCH AT RANK 1, HIGH SCORE



SUSPECT PHOTO
NOT NIJEER PARKS

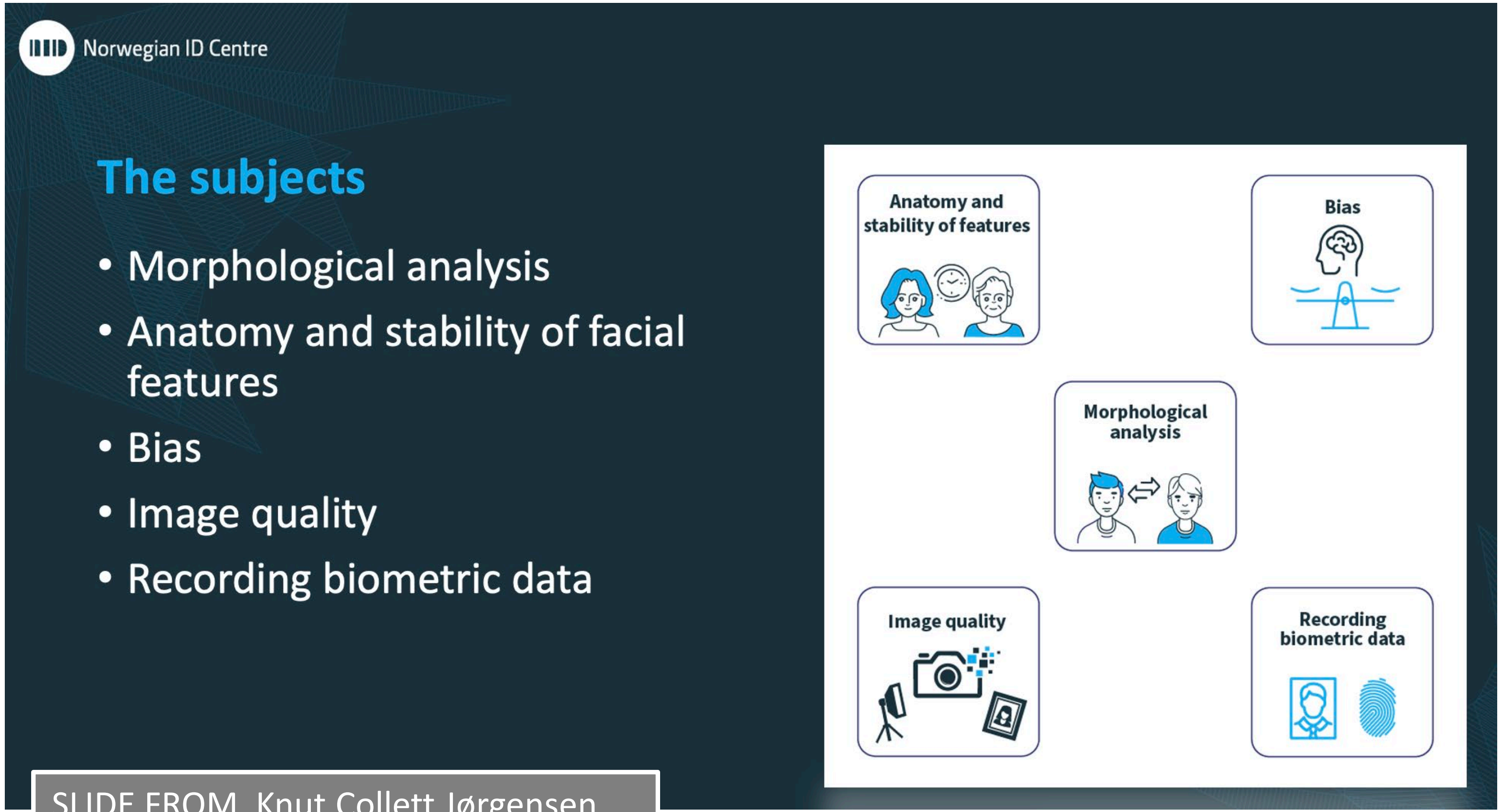
- 5 ALGS FIND GALLERY MATCH AT RANK 1, WEAK SCORE
- 1 ALG FINDS GALLERY MATCH AT RANK 8, WEAK SCORE

BLACK BOX STUDY OF HUMAN CAPABILITY :: VARIABLE AND POOR

- Examiners, reviewers, super-recognizers statistically the same
- Each subject group has large range of accuracy
- Four groups had subjects with AUC=1

P. J. Philipps et al. *Face Recognition Accuracy of Forensic Examiners, Super-recognizers, and Machines*
Proceedings of the National Academy of Sciences, May 2018

HUMAN REVIEW: TRAINING OFFERED BY NORWAY (BORDER) POLICE



SLIDE FROM Knut Collett Jørgensen
National Police Directorate, Norway

THREE ONLINE TRAINING PROGRAMS

1. Face Comparison
2. MAD
3. PAD (future)

https://www.nidsenter.no/face

Announcing new closed-box study for facial examiners

Administered by NIST

- jonathon.phillips@nist.gov
- amy.yates@nist.gov

Organized and conducted by NIST and U. of Texas at Dallas

J. Stoughton/NIST

- Perform detailed comparisons of faces in images
- Write detailed reports
- Prepared to testify in court
- Extensive training (2-4 years)

Current study: Cross-race closed-box study

Measure the accuracies of facial examiners comparing face images of Black and White individuals

Now recruiting!

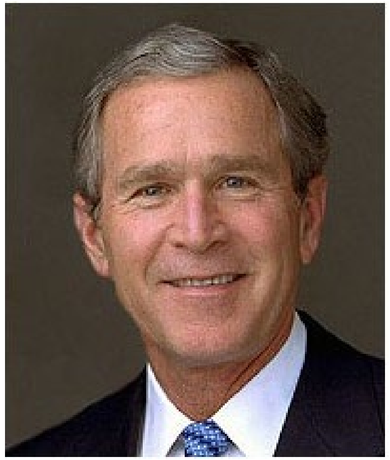
- Facial examiners
- Facial comparison professionals
- Super-recognizers

Email the organizers:

Face Recognition Under Attack: *Morphing*

WHO IS THIS?

FACE MORPHING: SINGLE IMAGE OF TWO PEOPLE



George W. Bush
(43rd US president)

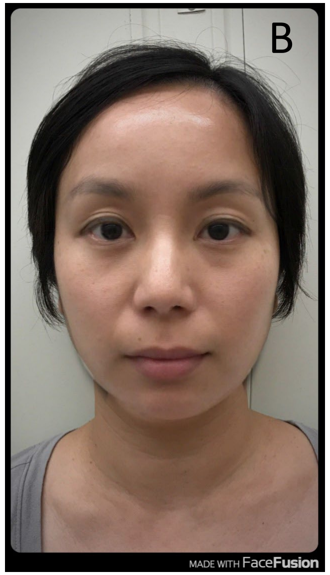
+

Barack Obama
(44th US president)

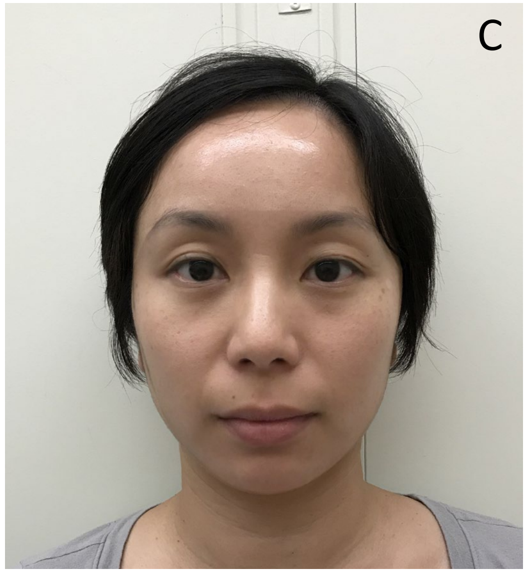
Face morphing generates an image that resembles both contributing subjects

MANY MORPHING TOOLS AVAILABLE (AND EVOLVING)

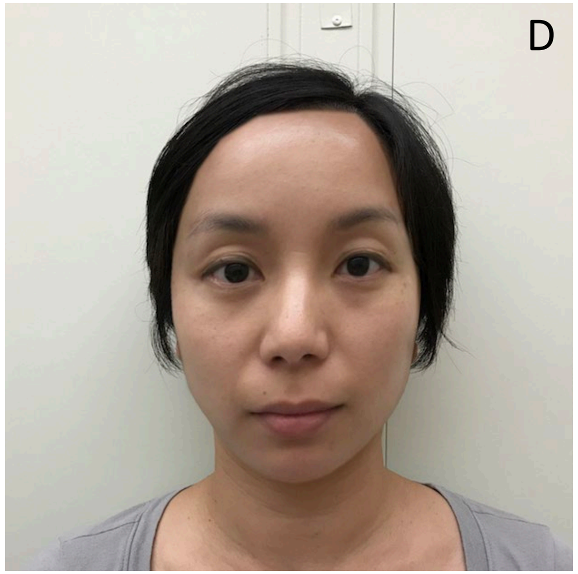
www.MorphThing.com



FaceFusion Mobile App

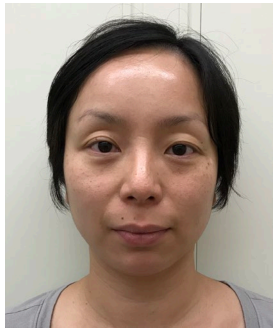


Automated Method



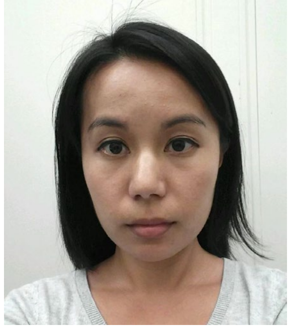
FantaMorph + Photoshop

Print + Scanned

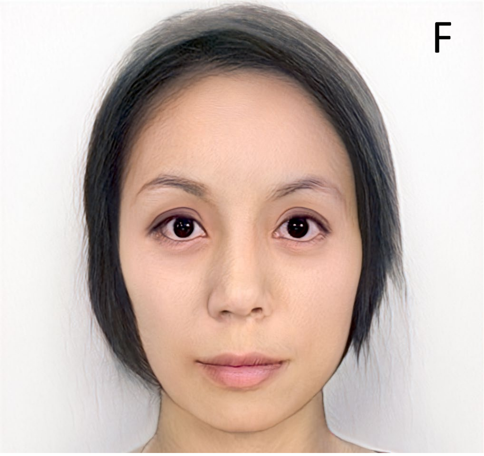


Subject A

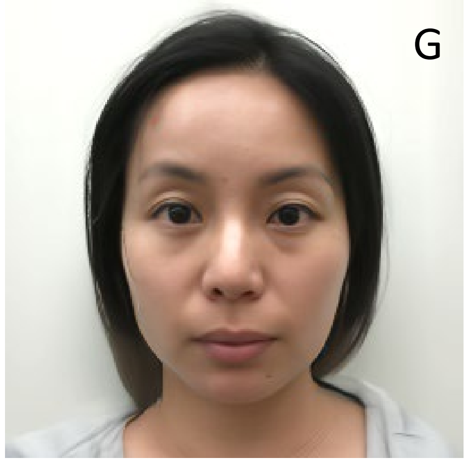
+



Subject B

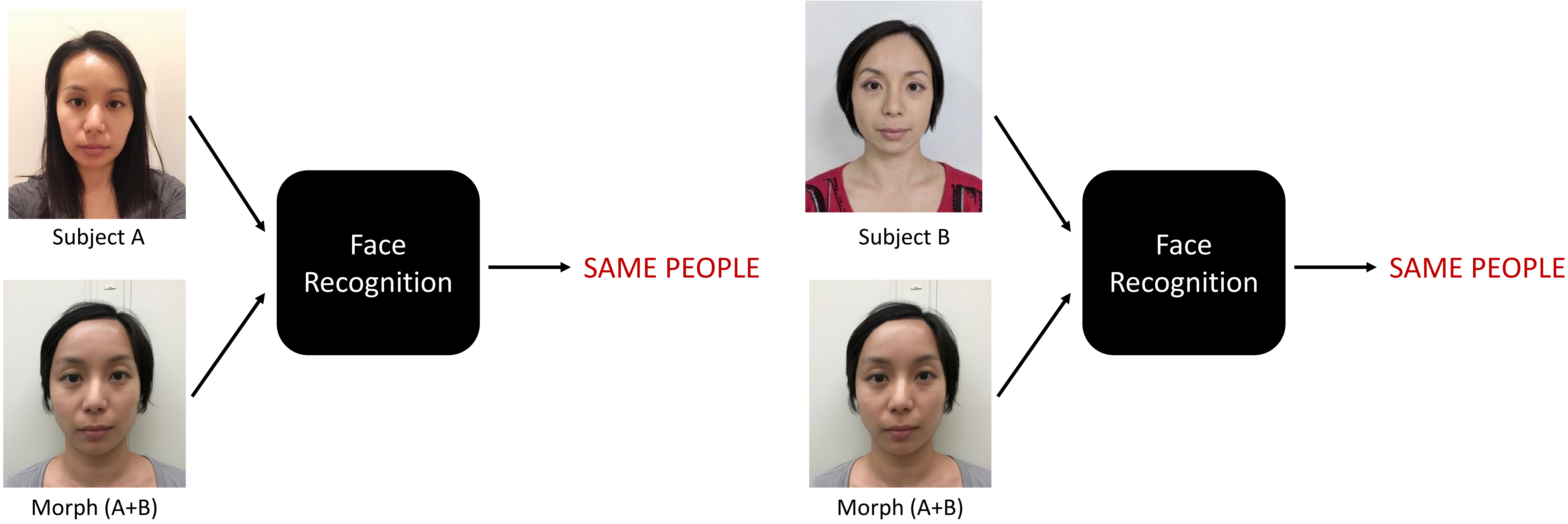


Generative Adversarial Network (GAN)



Diffusion Model

THE PROBLEM: FACE RECOGNITION MATCHES *BOTH* PERSONS



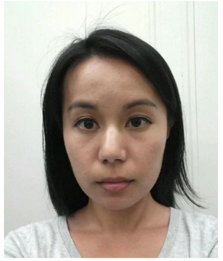
Multiple people can authenticate against a morph

All modern face recognition algorithms tested by NIST and operational matchers tested by parts of the U.S. Government are vulnerable to morphs

Image Source: NIST

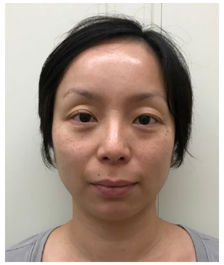
THREAT: ONE DOCUMENT, MULTIPLE USERS

PASSPORT APPLICATION



Accomplice
(passport owner)

+



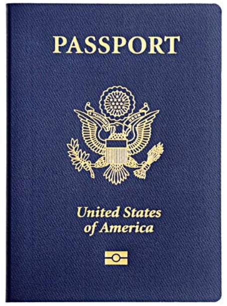
Attacker
(other identity)

Morph = A + B

Printed on
2in x 2in
photo paper and
mailed to passport
office

Passport office scans printed photo, conducts human examination, …

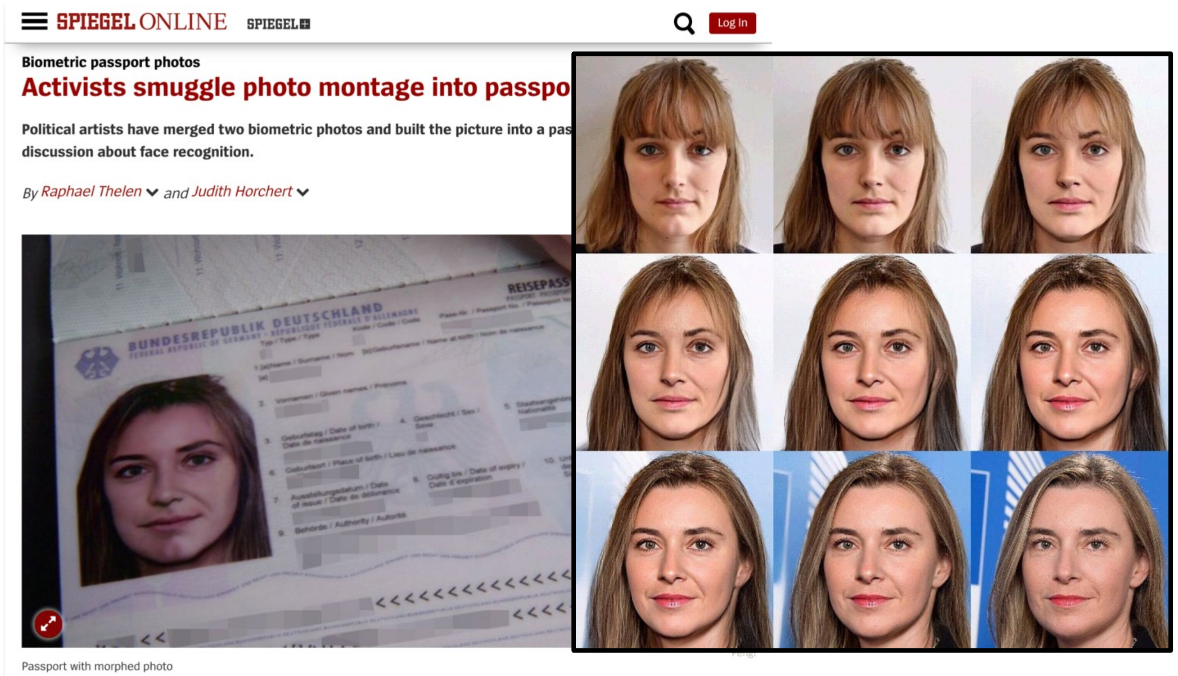
Ultimately, passport gets issued with morphed photo on it



Current U.S. passport application susceptible to manipulation of user-submitted photos.

Many other countries also accept user-submitted photos for identity credential applications.

REAL CASES OF MORPHING



Sept. 22, 2018: Member of German activist group successfully applies for a passport with a morphed image (containing Federica Mogherini, High Representative of the Union for Foreign Affairs and Security Policy)

Source (9/22/2018): http://www.spiegel.de/netzwelt/netzpolitik/biometrie-im-reisepass-peng-kollektiv-schmuggelt-fotomontage-in-ausweis-a-1229418.html via Google Translate

OVER 1000+ MORPHING CASES REPORTED ACROSS THE EU

Source: Presentation by Christoph Busch, Professor at NTNU/Hochschule Darmstadt at the International Face Performance Conference (IFPC) 2020, October 30, 2020 https://www.nist.gov/itl/iad/ig/ifpc-2020-conference-presentations-and-videos

SINCE 2020, OVER 40 MORPHING CASES WERE DETECTED IN SLOVENIA

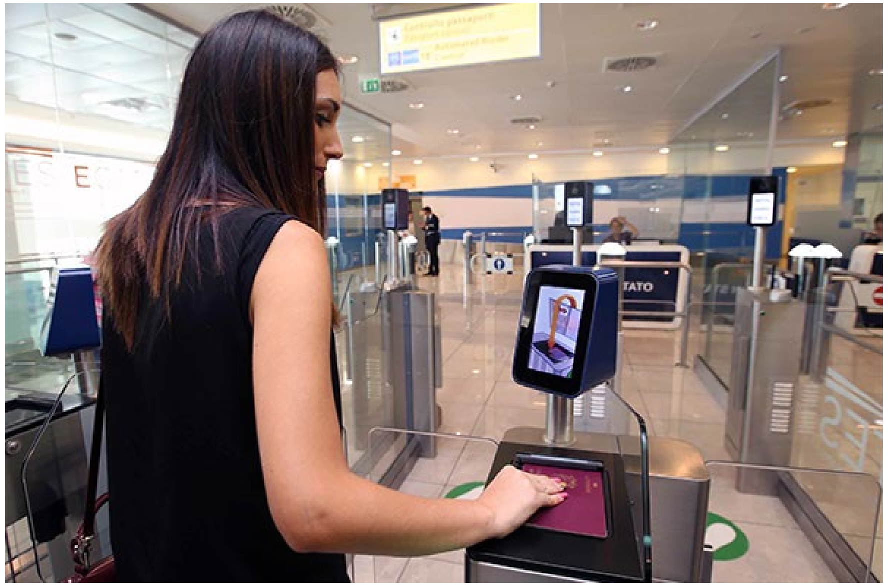
Source: Presentation by Matjaz Torkar, Deputy Commander of Station, Airport Police Station Brnik Slovenia at the International Face Performance Conference (IFPC) 2022, November 17, 2022 https://www.nist.gov/itl/iad/ifpc-2022-conference-presentations-and-videos

MORPHING: TWO DETECTION OPPORTUNITIES

Morph

Image Source: NIST

DOCUMENT ISSUANCE: Suspect image in isolation
Challenge: Morph detection is difficult and does not generalize across different/unseen morphing methods ⟶ impossible when attacker covers tracks.
Solution: Trusted photo capture



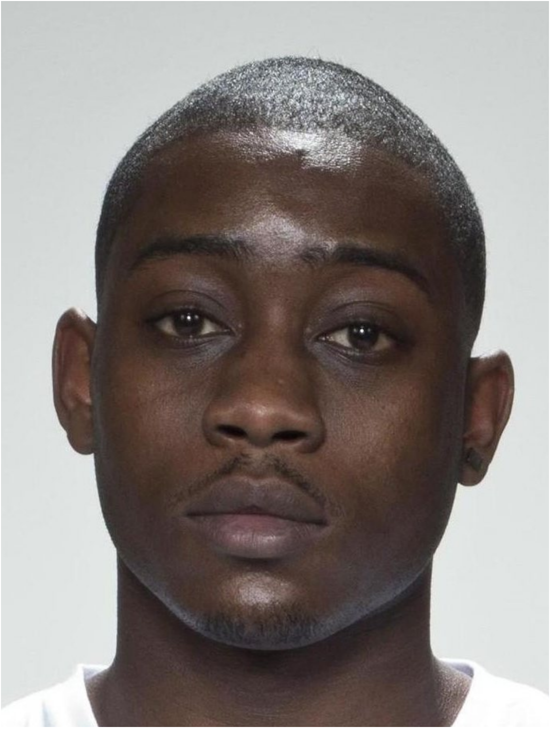
BORDER CROSSING: Suspect image + live image
Opportunity: Morph detection is possible (and more generalizable) because identity info can be analyzed (instead of specific image artifacts).

Image Source: http://www.futuretravelexperience.com/2016/01/automated-border-control-e-gates-go-live-at-naples-airport/

CAN HUMANS DETECT MORPHS?

A

B



C

Image Source: DeBruine, Lisa; Jones, Benedict (2017). Face Research Lab London Set. https://doi.org/10.6084/m9.figshare.5047666.v5

CAN HUMANS DETECT MORPHS?

A

MORPH

B

NOT A MORPH

C

NOT A MORPH

Image Source: DeBruine, Lisa; Jones, Benedict (2017). Face Research Lab London Set. https://doi.org/10.6084/m9.figshare.5047666.v5

Detecting morphed / manipulated face images*

In progress: MAD *(Morphing Attack Detection)*

- Developing training in detecting morphed face images

Next step: PAD *(Presentation Attack Detection)*

- Developing training in detecting otherwise manipulated face images.
  - E.g. geometric changes (barrel/pincushion distortion, manual unintended (?) changes..)

- Will be available on: www.nidsenter.no/face *-NB: Under construction*

- * Based on findings from the iMARS project (https://imars-project.eu)

SLIDE FROM Frøy Løvåsdal
Team Identity, biometrics and biometric data EUIS-programme
National Police Directorate, Norway
froy.lovasdal@politiet.no

MORPHING: POSSIBLE MITIGATION

NIST

Do live enrollment

- Norway (now), Sweden (now), Germany 2025[1]
- Should be adopted by all countries to be effective
- But some morphs in circulation now

Eliminate print + scanned photos

- Avoid printing and scanning
- Require high resolution, digital photos

Use FR on centralized database

- Perform 1:N duplicate check; look for suspicious activity [NISTIR 8430]

Do trusted external capture

- Signed photobooths
- Certified photographers (e.g., Finland, France)
- Liveness detection in dedicated, secure mobile application

Build awareness

- Train relevant personnel about morphs
- Can training improve personnel skills on morphed image over time?
- What cues are people good at detecting morphs using and are any of them tangible to document?

Establish strong secondary verification processes

- Verify with additional data source (e.g., Slovenia)
- Use another biometric modality

[1] https://www.reuters.com/article/us-germany-tech-morphing/germany-bans-digital-doppelganger-passport-photos-idUSKBN23A1YM

Face Recognition Under Attack: *Presentation Attack*

"the presentation of an artefact or of human characteristics to a biometric capture subsystem in a fashion intended to interfere with system policy"."

Source: JTC1/SC37 (2023) International Organization for Standardization: Information Technology – Biometric presentation attack detection – Part 1: Framework. ISO/IEC 30107- 1

UNATTENDED FR IS OPEN TO PRESENTATION ATTACK

"… passenger who boarded plane in Hong Kong as an old man in flat cap and arrived in Canada a young Asian refugee"

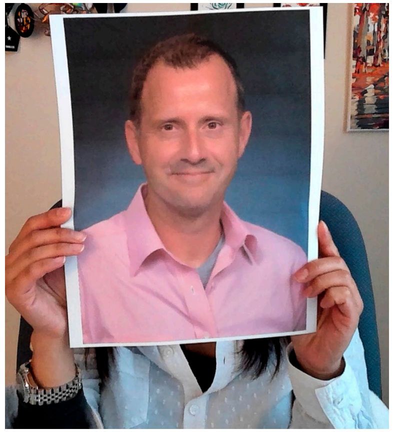
http://www.dailymail.co.uk/news/article-1326885/Man-boards-plane-disguised-old-man-arrested-arrival-Canada.html

https://adam.harvey.studio/cvdazzle

1. Spoofing / Impersonation:

- Goal: Impersonate someone else
- How: Maximize similarity score

Enrollment



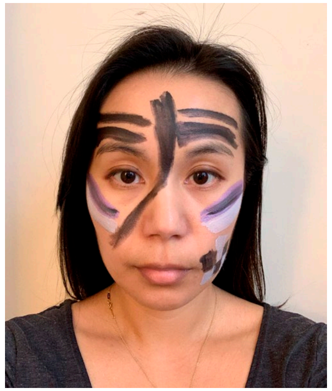
Verify

Verify

2. Evasion:

- Goal: Do not match your prior enrollment, to impede 1:N detection
- How: Minimize similarity score

Database

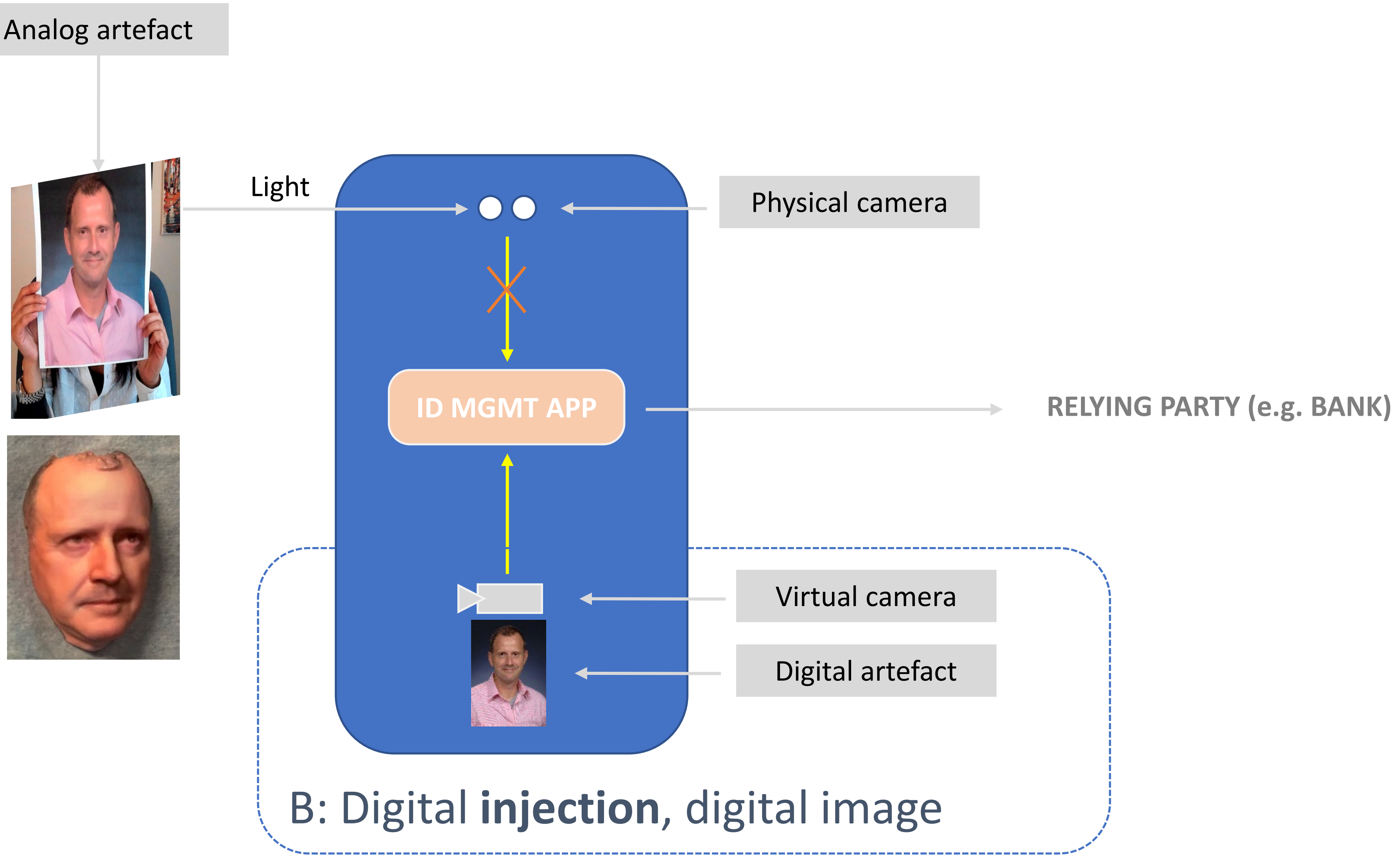


Search

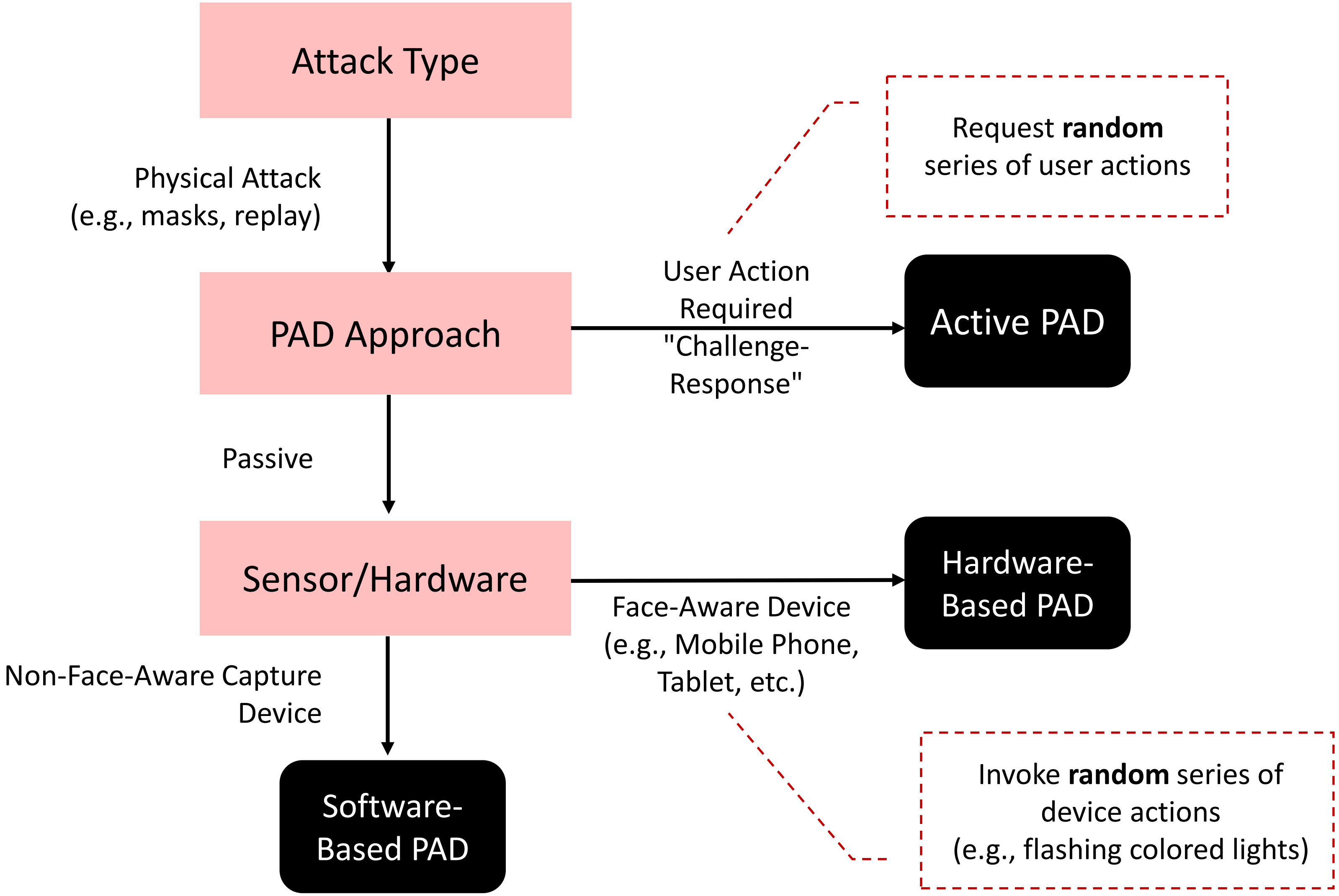
ANALOG VS. DIGITAL ATTACKS

A: **Physical attack**, analog artefact

- Rubber mask
- Printed photo
- Tablet display

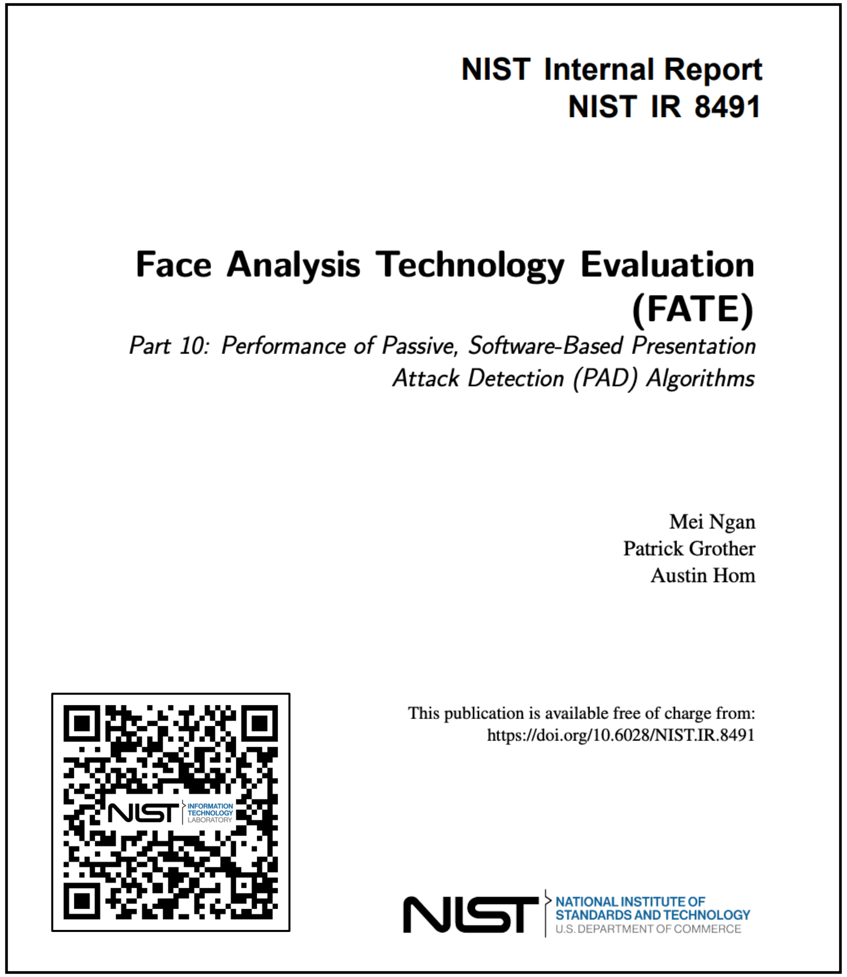


APPROACHES TO PRESENTATION ATTACK DETECTION (PAD)



NIST's PAD BENCHMARK

PAD at NIST


NIST Internal Report
NIST IR 8491

Face Analysis Technology Evaluation (FATE)

Part 10: Performance of Passive, Software-Based Presentation Attack Detection (PAD) Algorithms

Mei Ngan
Patrick Grother
Austin Hom

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8491

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY U.S. DEPARTMENT OF COMMERCE

Attack Type

Physical Attack
(e.g., masks, replay)

PAD Approach

User Action Required → Active PAD

Passive

Sensor/Hardware

Face-Aware Device
(e.g., Mobile Phone, Tablet, etc.) → Hardware-Based PAD

Non-Face-Aware Capture Device → Software-Based PAD

PAD at DHS

Remote Identity Validation Technology Demonstration 3

NIST's PAD BENCHMARK

- Scope: evaluate **passive**, **software-based PAD** (with still photograph(s) and video frames)
- Application
  - Server-based or cloud-based PAD with non-face-aware capture device
  - Offline PAD in existing/legacy systems
- Methodology
  - Tested two separate use cases – ability to detect
    - Impersonation
    - Evasion/Concealment
  - Evaluated **82** algorithms from **45** unique developers worldwide (2 month submission window)
  - Ran on attack images of various species (9 PA types)
- Results
  - NISTIR 8491 report published Sept. 20, 2023

+ http://www.dailymail.co.uk/news/article-1326885/Man-boards-plane-disguised-old-man-arrested-arrival-Canada.html

SOFTWARE-BASED PAD – WHAT WE FOUND

- PAD performance varied significantly across algorithms, use cases, and attack types

- THE GOOD NEWS
  - The detection **photo print/replay attacks, protective face masks, and flexible silicon face masks** was well supported

  - **Fusion** of multiple PAD algorithms improved accuracy

  - Higher accuracy in **video** sequences vs. single image

- THE BAD NEWS
  - No algorithm worked well at detecting all attack types

  - There remain PA types for which detection error rates are high (we did not disclose which types)

Iris Recognition

Most recent adoption of iris recognition

https://abcnews.go.com/Business/apple-vision-pro-cost-3499-people-pay/story?id=106509013

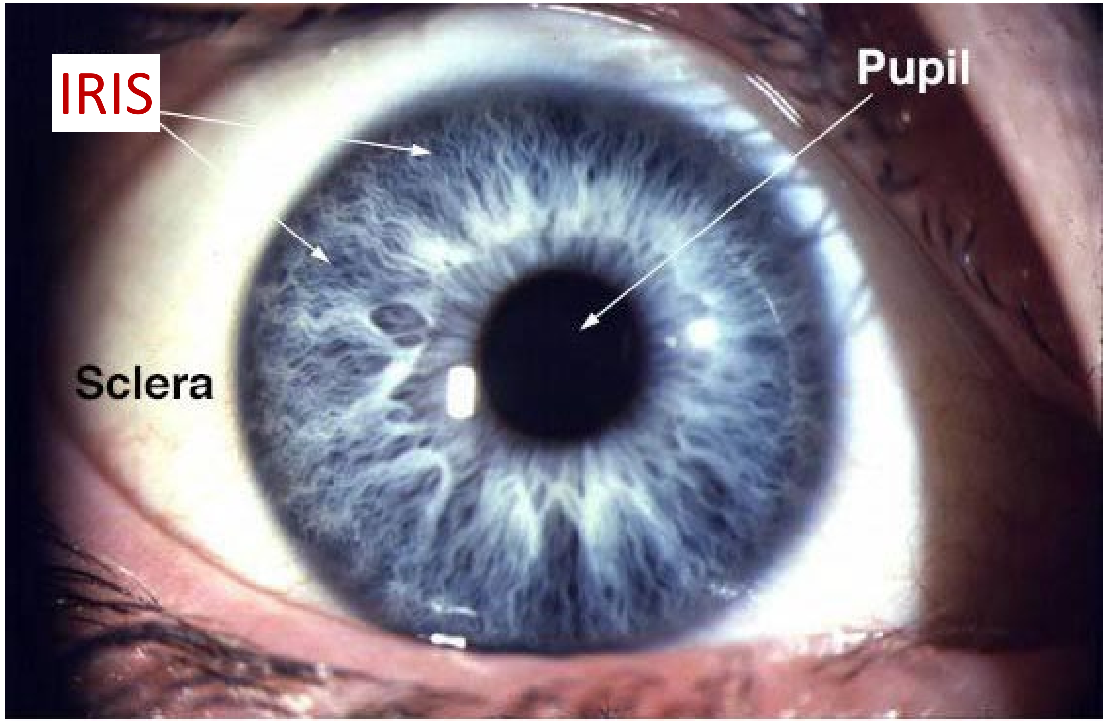
The probability that a random person in the population could unlock your Apple Vision Pro using Optic ID is less than 1 in 1,000,000.

…

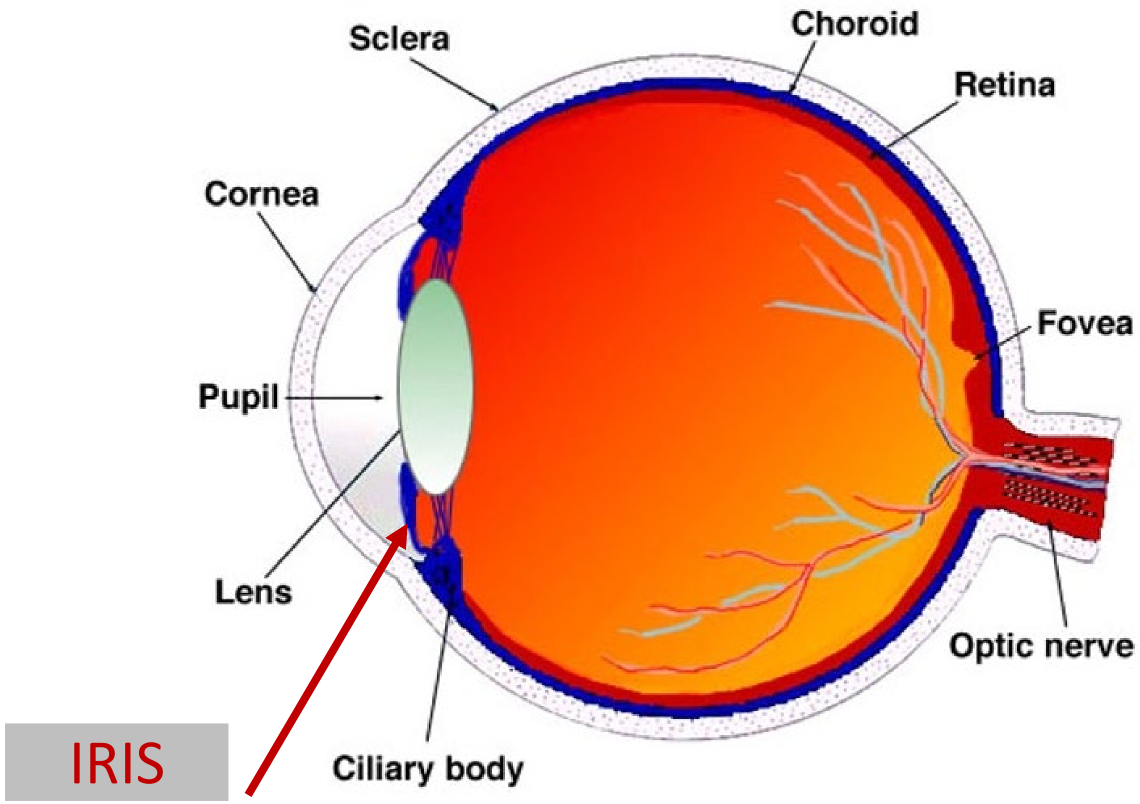
Optic ID matches against detailed iris structure in the near-infrared domain, which reveals highly unique patterns independent of iris pigmentation. It's designed to protect against spoofing through the use of sophisticated neural networks that analyze the authenticity of the iris and surrounding region.

https://support.apple.com/en-us/118483
Retrieved 2024-04-22

Iris Anatomy



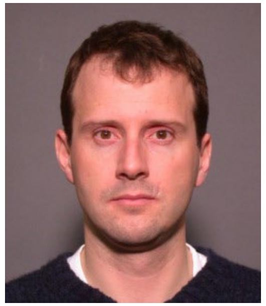
Gross Anatomy of the Eye, Helga Kolb, 2005
https://www.ncbi.nlm.nih.gov/books/NBK11534/



Iris is NOT retina!

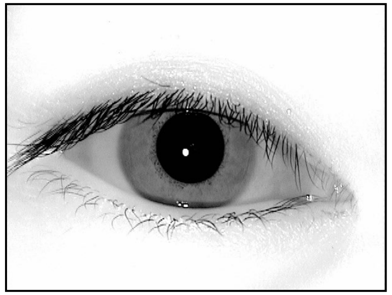
Face vs. Iris

Face



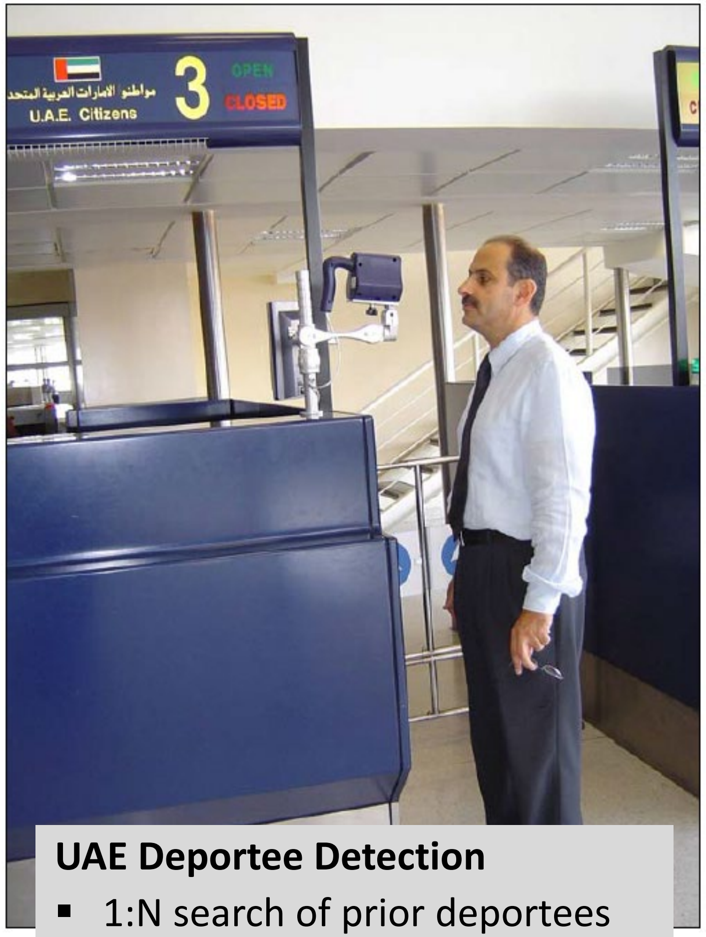
- Photography
  - Frontal portrait
  - Visible light
  - Commodity camera
  - Exch. standard: 39794-5:2019
  - Qual. standard: 29794-5
- Enrollment
  - Best image size: ~1600 x 1200
  - Minimum image size: 640 x 480
- Uniqueness: No; Limited by twins, siblings
- Permanence: Good
- Impersonation: Easier to collect sample
- Human adjudication:
  - Training required; aptitude varies
  - Human bias
- Passport LDS: DG2, 15KB image

Iris



- Photography
  - Two eyes simultaneously
  - **Infrared light**
  - Specialized camera
  - Exch. standard: 19794-6:2011
  - Qual. standard: 29794-6
- Enrollment
  - Best image size: 640 x 480
  - Minimum image size: 400 x 400
- Uniqueness: Yes; twins do not give false positives
- Permanence: Fair
- Impersonation: Harder to collect sample covertly
- Human adjudication:
  - Training required
- Passport LDS: DG4, 3KB image

Deployments



UAE Deportee Detection
- 1:N search of prior deportees
- Non-citizens
- Operational since 2003

Aadhaar India: National ID
- 1:N duplicate detection using 2 iris + 10 fingers (+ face)
- 1.2B residents 2018-07
- Operational since 2010-09

Singapore Ports
- Face + Iris + Fingerprint

https://www.ica.gov.sg/news-and-publications/newsroom/media-release/use-of-iris-and-facial-biometrics-as-the-primary-biometric-identifiers-for-immigration-clearance-at-all-checkpoints

Source: Sky News
http://www.youtube.com/watch?v=51Num5h7itk

http://uidai.gov.in/images/FrontPageUpdates/role_of_biometric_technology_in_aadhaar_jan21_2012.pdf

NIST's IREX Leaderboard

	Matcher	Submission Date	Accuracy (FNIR)	Search Time (sec)	Template Creation Time (sec)	Template Size (bytes)	FTE Rate
1	NEC	Dec 2022	0.0022 ± 0.0004	12 ± 3	1.03 ± 0.06	17 374 ± 2 729	0
2	Innovatrics	Apr 2023	0.0024 ± 0.0003	3.2 ± 0.6	1 ± 1	8 277 ± 115	0.000007
3	Idemia	Jun 2023	0.0026 ± 0.0004	11 ± 5	1.5 ± 0.1	129 206 ± 6 019	0
4	Hikvision	Jan 2023	0.0029 ± 0.0004	16 ± 7	3 ± 1	15 404 ± 0	0
5	Neurotechnology	Dec 2023	0.0029 ± 0.0004	32.3 ± 0.6	0.5 ± 0.2	25 788 ± 0	0
6	Thales	Dec 2022	0.0030 ± 0.0004	14 ± 6	1.6 ± 0.6	43 362 ± 0	0
7	IrisID	Feb 2023	0.0034 ± 0.0004	2.1 ± 0.5	0.16 ± 0.01	5 636 ± 0	0.0001
8	Irlinker	Oct 2023	0.0044 ± 0.0005	17.4 ± 0.2	1.18 ± 0.02	28 159 ± 133	0.0003
9	EyeCool	Jan 2023	0.0044 ± 0.0005	84 ± 48	0.422 ± 0.007	63 684 ± 0	0
10	Dermalog	Feb 2023	0.0048 ± 0.0005	1.88 ± 0.05	0.73 ± 0.03	3 915 ± 39	0
11	Decatur	Nov 2021	0.0060 ± 0.0005	32 ± 2	1.4 ± 0.2	40 096 ± 6 427	0
12	ROC	Oct 2023	0.0072 ± 0.0006	0.117 ± 0.008	0.581 ± 0.007	528 ± 0	0

Snapshot: 2024-05-06

https://pages.nist.gov/IREX10/

IRIS EXCHANGE

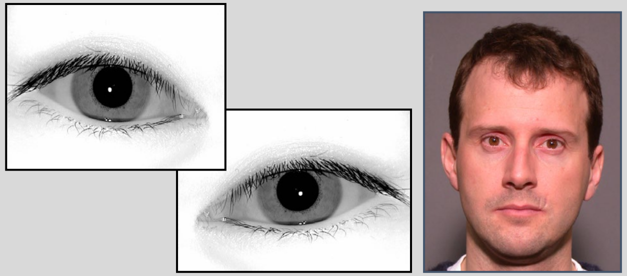
FACE + IRIS = A COMBINED MODALITY

https://www.idemia.com/walk-through-multi-biometric-solution

https://www.irisid.com/productssolutions/hardwareproducts/icam-d2000/

https://www.nec.com/en/press/202211/global_20221108_01.html

https://cmi-tech.com/product/ef-45nc-dual-iris-recognition-system/



MULTIMODAL PACKET

MULTIMODAL OPTIONS

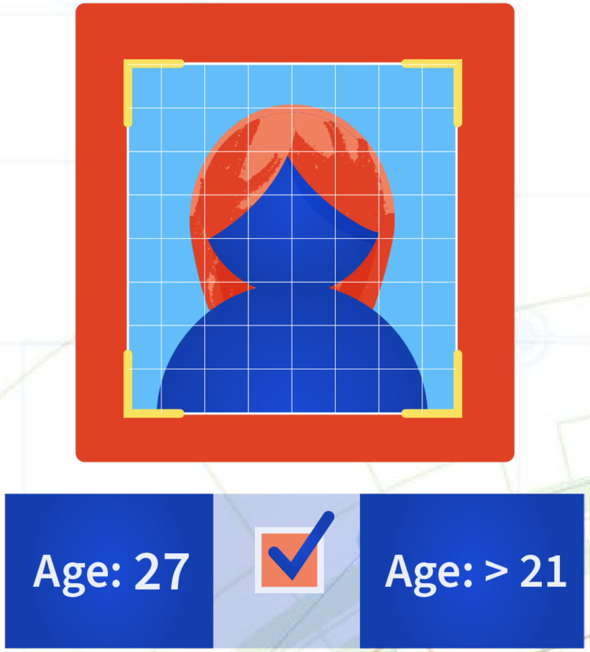
- “EITHER-OR” ⟶ Low FNIR
- “BOTH” ⟶ Very low FPIR, larger N
- "BOTH" ⟶ Presentation attack is more difficult

- Demographic differences ⟶ Reduced
- Twins[1,2]

[1] J. Daugman, "How iris recognition works," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 21-30, Jan. 2004.
[2] Zhenan Sun, Alessandra A. Paulino, Jianjiang Feng, Zhenhua Chai, Tieniu Tan, Anil K. Jain, "A study of multibiometric traits of identical twins," Proc. SPIE 7667, Biometric Technology for Human Identification VII, 76670T (14 April 2010);

Q&A (10 minutes)

Age Estimation & Verification



APPLICATIONS OF AGE ESTIMATION & VERIFICATION (AEV)

NIST

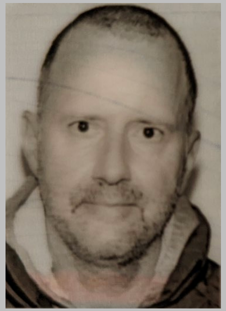
- ACTIVE
  - **AGE RESTRICTION SALES:** Is person old enough e.g., 18 for cigarettes?
  - **ONLINE SAFETY:** Is person within an age range e.g., 13-16 online chat room
  - **BORDER CONTROL:** How old is this refugee, asylum seeker … ?
- PASSIVE
  - **ADVERTIZING:** Age-tailored digital display ads
  - **INSIGHT:** Age statistics for people in certain locations (e.g., movie theaters)
- OTHERS

NIST AGE ESTIMATION + VERIFICATION BENCHMARK
THREE AEV FUNCTIONS

#1: ESTIMATE_AGE FROM SINGLE IMAGE

⟶ AGE = 36.4

#2: VERIFY_AGE > 25 FROM SINGLE IMAGE



⟶ TRUE

#3: ESTIMATE_AGE FROM

+

⟶ AGE = 36.4

1. Legislative action
2. Applications
3. Three functions, spur innovation

4. 2023-08-14: Published v1 API
5. 2023-09-05: Open to algorithms
6. 2024-05: First report

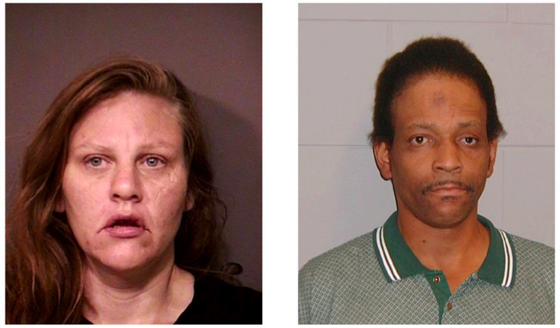
7. Report
    - Gains
    - Accuracy
    - Speed
    - Demographic dependence
    - Quality dependence
    - Eye glasses

8. Standards
    - ISO/IEC 27566
    - IEEE 2089

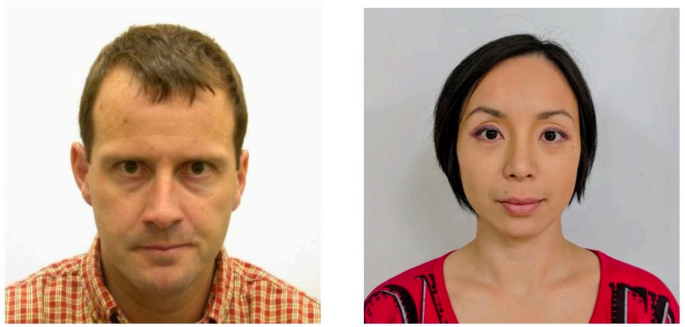
BEYOND SCOPE: AGE ASSURANCE ⟶ TRUSTED AV including LIVENESS + PAD + FALLBACKS

FATE AEV: DATASETS

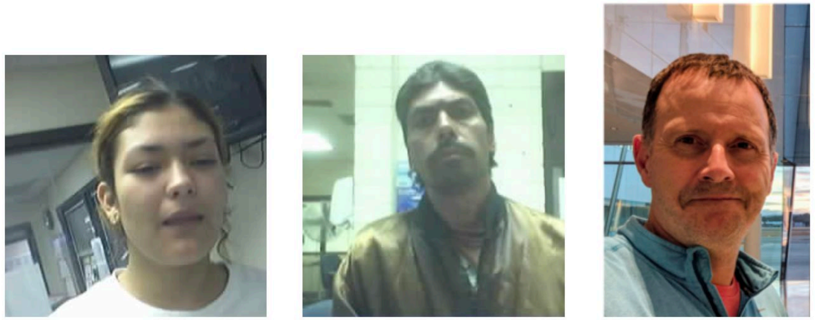
Name	Sec	Age Precision	Num. Images	Num. People	Purpose
Visa	3.1	Day	6249294	5738091	Exact repeat of 2014 study
Mugshots	3.2	Year	1482667	1482667	AE accuracy on standardized photos
Application	3.3	Day	1054704	802332	Challenge-T and demographics
Border	3.4	Day	2715230	632520	Analysis of effect of quality
Kalina Everyday	3.5	Day	1991	1	Longitudinal ageing



Mugshot Photos



Application Photos



Border Photos

AEV ACCURACY GAINS SINCE 2014

NIST

Incode's age estimate is 1.19 years better, on average, than Cognitec 2014

Algorithm	MAE
incode-000	3.08
yoti-001	3.30
unissey-001	3.87
roc-000	3.81
dermalog-001	4.01
cognitec2013Oct-001	4.27
neurotechnology-000	4.54
nec2013Oct-002	5.32

Gains since 2014: Age estimation error statistics for the 2024 algorithms and the two most accurate from 2014. The value mean absolutely error (MAE). Dataset is the Mexican visa population.

APPLICATION/METRIC: "CHALLENGE – 25" AGE RESTRICTED SALES

NIST

Legal Age (LA)

Challenge Age (T)

BUFFER

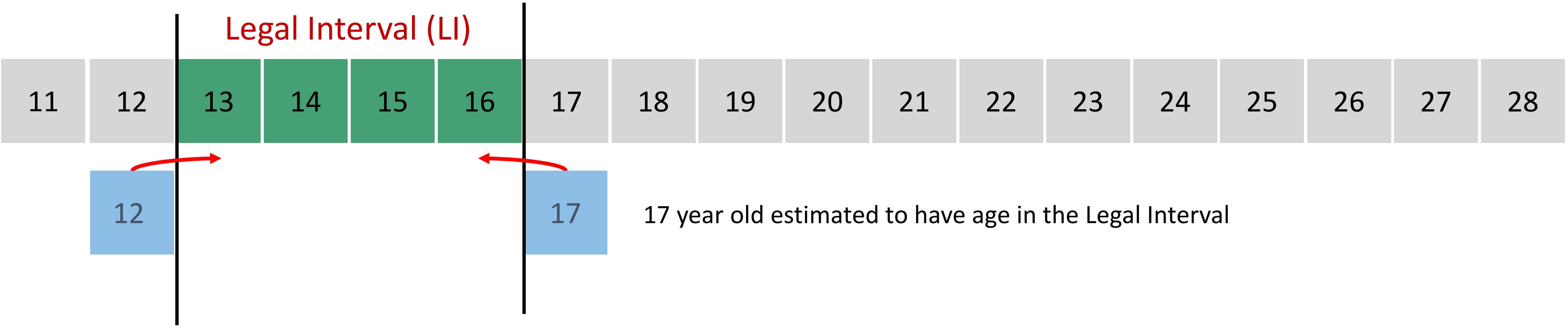
12	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28

17	17 year old estimated to have age 25 or above	≥ 25	FPR_{17}
16	16 year old estimated to have age 25 or above	≥ 25	FPR_{16}

https://www.biometricupdate.com/202302/customers-prefer-facial-age-estimation-for-self-checkouts-yoti-global-retail-report

Policy: Anybody who is classified as under LA=25 is challenged to prove age another way (photo ID etc.)

APPLICATION/METRIC: KIDS ONLINE SAFETY



CURRENT DATASETS

1. TWO GLOBAL SETS WITH 14 - 99
2. SINGLE-COUNTRY SET WITH 0-99

FUTURE (??)

3. ONE GLOBAL SET WITH 0-99

IMAGE QUALITY MATTERS

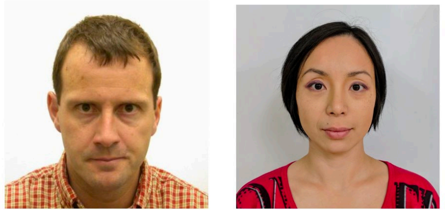
NIST

Challenge-25 False Positive Rate for subjects aged 17 (Application vs. Border Photos). Lower values are better.

APPLICATION PHOTOS

MALE

Age	E. Africa	E. Asia	E. Europe	S. Asia	SE. Asia	W. Africa
17	0.02	0.006	0.003	0.017	0.01	0.05



Application Photos

BORDER PHOTOS

MALE

Age	E. Africa	E. Asia	E. Europe	S. Asia	SE. Asia	W. Africa
17	0.43	0.038	0.12	0.15	0.05	0.36

Algorithm: incode-000

Border Photos

DEMOGRAPHICS MATTER

Dataset: Application images, Challenge-25 False Positive Rate for subjects aged 17 by sex and region of birth. Lower values are better.

Algorithm	Male E. Europe
dermalog-001	0.04
incode-000	0.003
neurotechnology-000	0.06
roc-000	0.00
unissey-001	0.04
yoti-001	0.003

REGION OF BIRTH MATTERS

Dataset: Application images, Challenge-25 False Positive Rate for subjects aged 17 by sex and region of birth. Lower values are better.

Algorithm		Male E. Africa	Male E. Asia	Male E. Europe
dermalog-001		0.10	0.07	0.04
incode-000		0.02	0.006	0.003
neurotechnology-000		0.76	0.36	0.06
roc-000		0.07	0.12	0.00
unissey-001		0.20	0.26	0.04
yoti-001		0.02	0.036	0.003

REGION OF BIRTH + SEX MATTERS

NIST

Dataset: Application images, Challenge-25 False Positive Rate for subjects aged 17 by sex and region of birth. Lower values are better.

Algorithm	Female E. Africa	Female E. Asia	Female E. Europe	Male E. Africa	Male E. Asia	Male E. Europe
dermalog-001	0.13	0.14	0.31	0.10	0.07	0.04
incode-000	0.11	0.05	0.07	0.02	0.006	0.003
neurotechnology-000	0.39	0.24	0.12	0.76	0.36	0.06
roc-000	0.18	0.17	0.02	0.07	0.12	0.00
unissey-001	0.17	0.26	0.19	0.20	0.26	0.04
yoti-001	0.18	0.14	0.19	0.02	0.036	0.003

AEV: SUMMARY AND THOUGHTS

NIST

Performance

- Accuracy has improved since 2014
  - Five of the six algorithms outperform the most accurate algorithm submitted in 2014
- Accuracy varies across algorithms
  - No single standout algorithm
  - Variation across image quality, sex, region of birth, subject age
- Operationally, presentation attack detection is usually required for active applications
  - Coupled to the AEV system
- Age estimation will never be perfect (but does it need to be?)

Next Steps

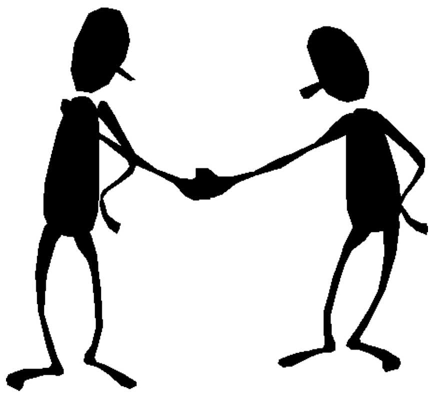
- AEV accuracy will continue to evolve (AE 2014 < AE 2024 < AE tomorrow)
  - Development continues
    - Noise suppression (via dataset augmentation?)
    - Bias correction: Women and non-Europeans (via diversification of training data)
- FATE AEV is an ongoing resource available to developers + purchasers + policy makers
  - FATE AEV is open to new developers and new algorithms
  - Also evaluate differential AE

Contactless Fingerprint Technology

CRADA

Cooperative Research and Development Agreement

Contactless Fingerprint Capture Device Measurement



https://www.nist.gov/itl/iad/image-group/crada-program-contactless-fingerprint-capture-device-measurement

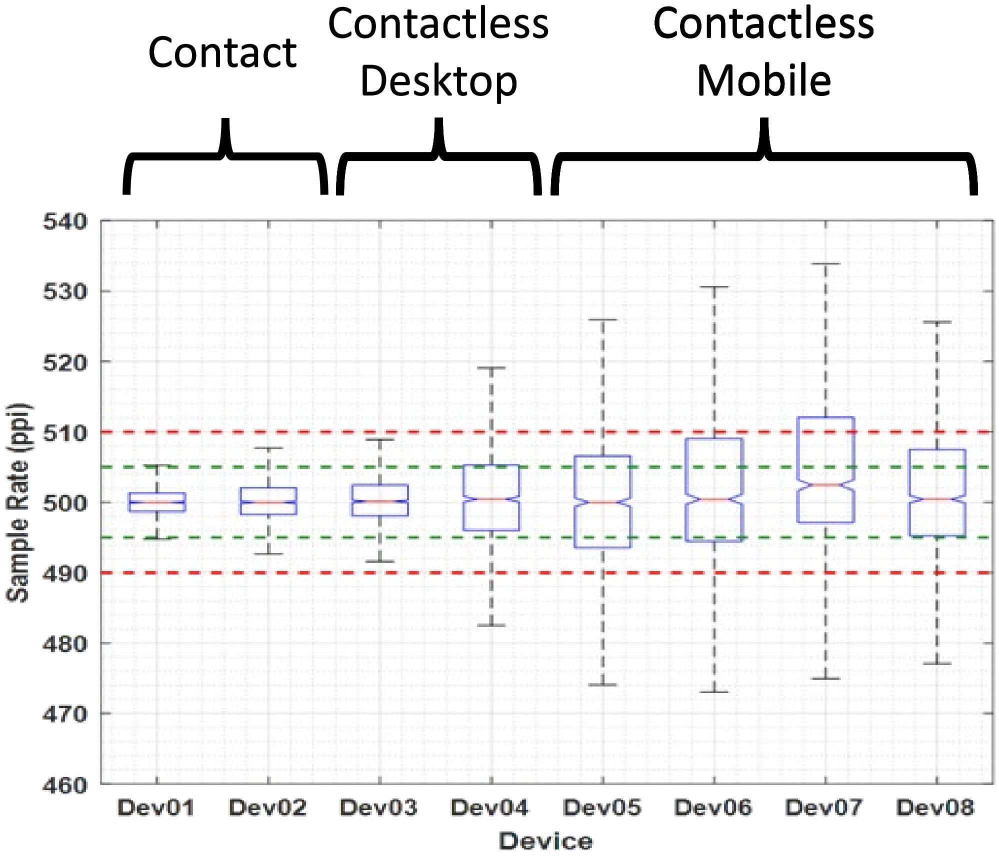
Fundamental Differences

NISTIR-8307

Some caused by poor capture controls

Since most contactless fingerprint collection devices utilize a photographic process, these devices share many of the same challenges.

For example, since distance may be poorly constrained, sample rate is not well controlled.



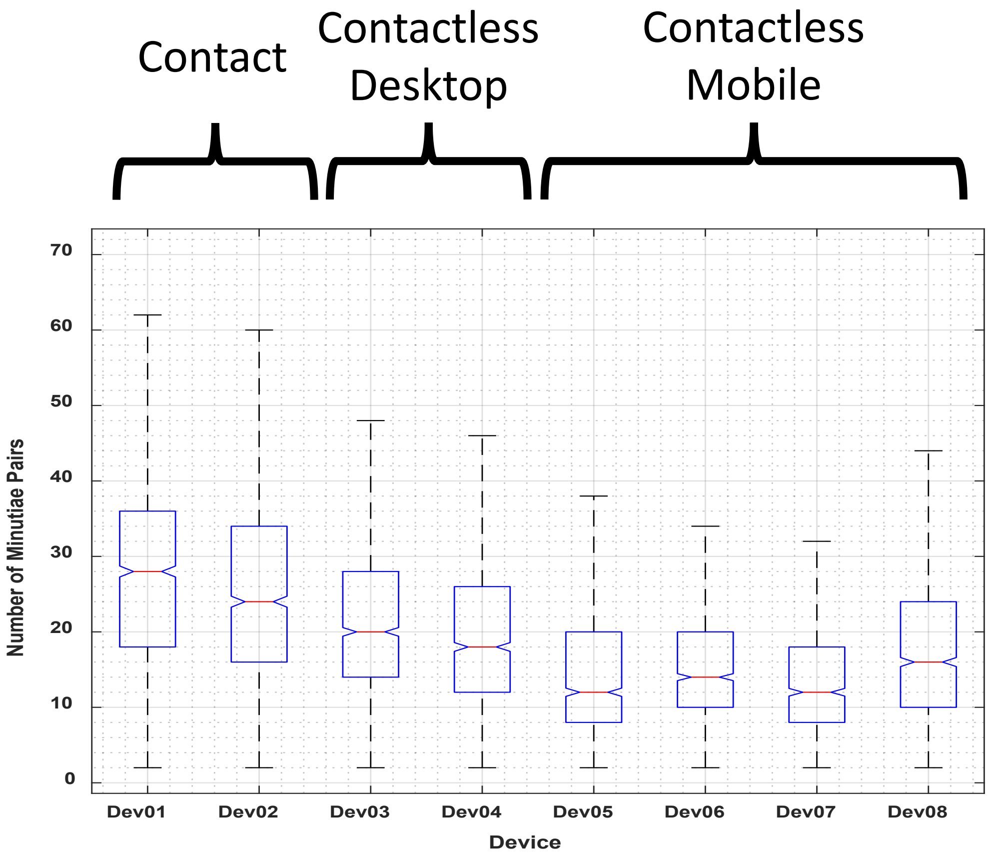
Sample Rate Differences

Fundamental Differences

Some caused by legacy algorithm limitations

At a signal level, the images differ sufficiently to render some of the algorithms developed for contact collected friction ridge imagery to not perform as well on contactless.

Results raised caution flags on the usage of existing/traditional algorithms without due diligence.



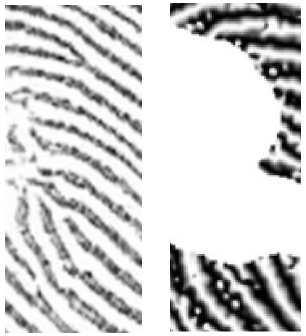
Feature Detection

Examples

Contact Contactless

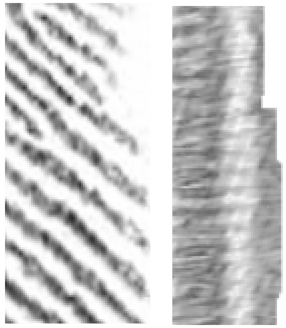
Too much detail

Contact Contactless



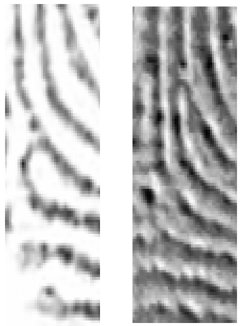
Not enough detail

Contact Contactless

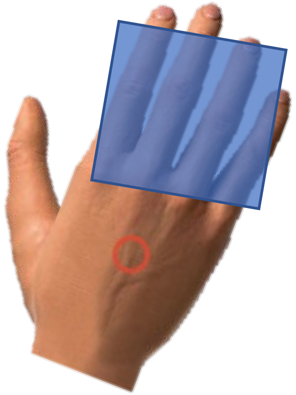


Depth of field

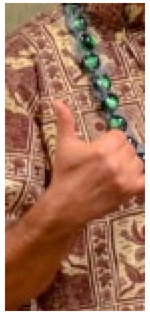
Contact Contactless



Contrast Inversion



Capture volume / aperture



Pose / Backdrop

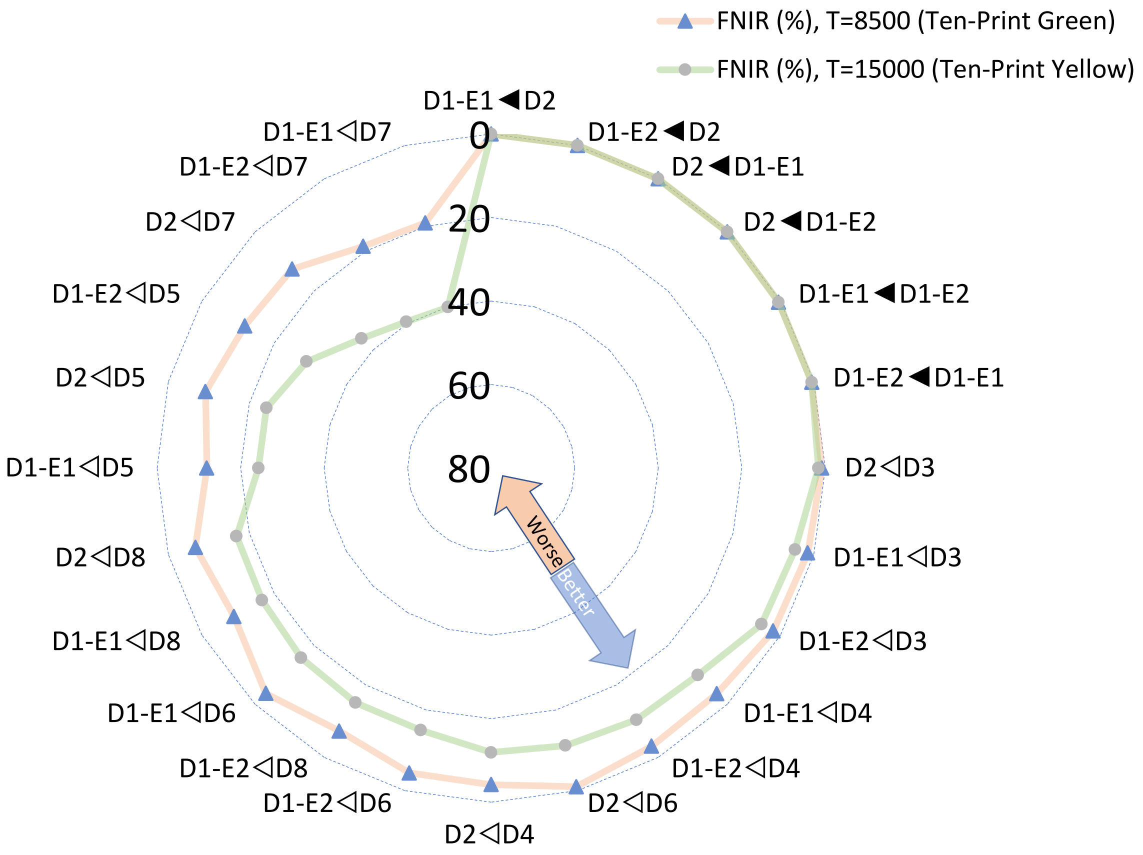
Sample images from NIST employee, used with permission.

Fundamental Differences

Ultimately impacts Machine Behavior

[NISTIR-8315]

All this wobble in the data incurs a significant penalty in accuracy, and reinforces the need for standardization at the first mile of data collection (capture)



Legend:	D1: Contact, FTIR	D2: Contact, EL	D3: Contactless, Desktop	D4: Contactless Desktop	"◀": Contact Cases Only
	D5: Contactless Mobile Phone	D6: Contactless Mobile Phone	D7: Contactless Mobile Phone	D8: Contactless Mobile Phone	"◁": Contactless Cases

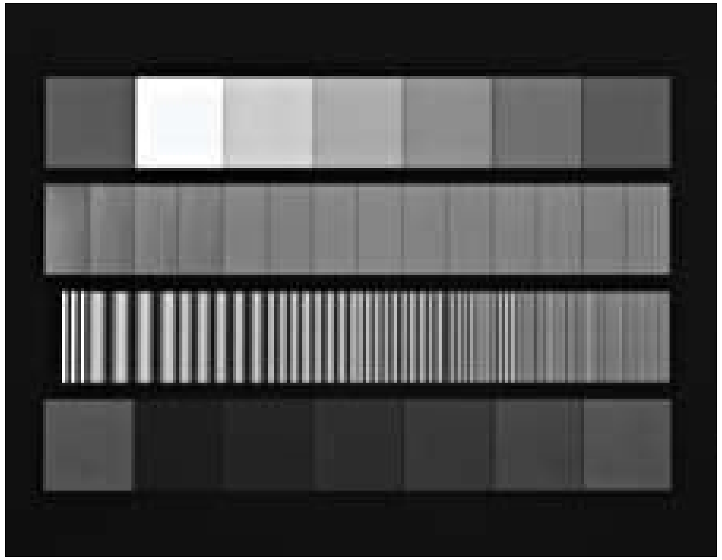
Calibration and Certification Guidance

Contact

Consistent Collection for Contact Devices: Appendix F

Legacy devices are well tested & certified, but the old tests may or may not work on these new devices.

i.e., Linearity/geometric accuracy/CTF/MTF (MTR 01B0000021)



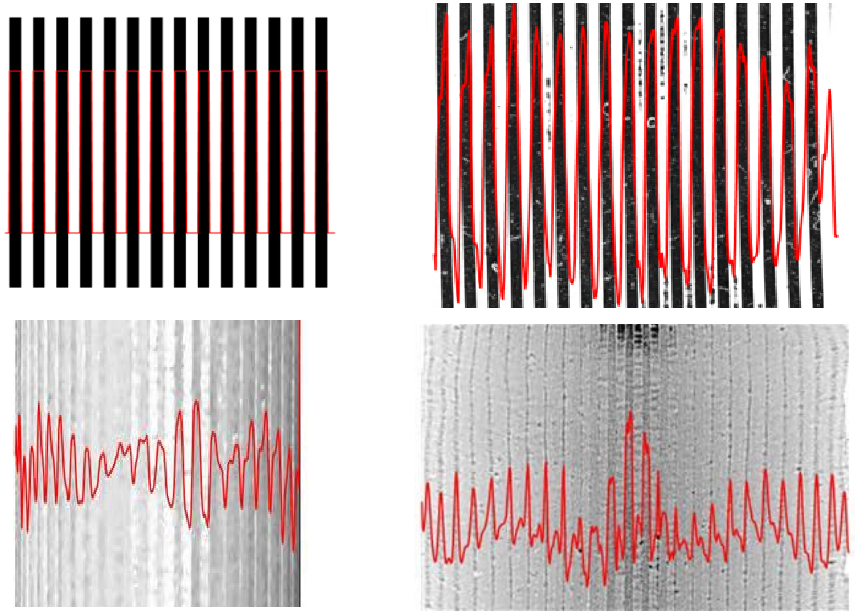
Tools and Targets

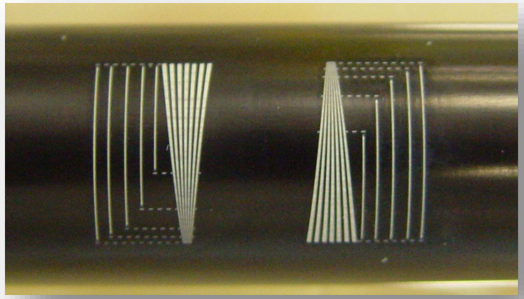
Targets: A known artifact… with known specs.

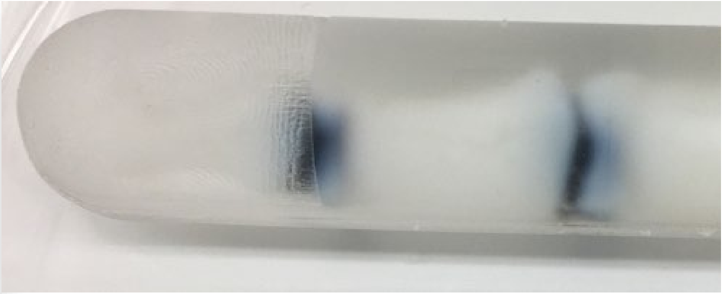
2010: Began developing calibration targets.

2016: Granted USG Patents.

2017+: On going evolution of test targets.







Target Work Continues

- Targets will continue to play an important role and will enhance testing as we go forward.
- Targets may be key for continuous calibration of devices.
- Targets may also be crucial in examining forensic fidelity of contactless, at scale.
- Targets may also be key to certification of newer exotic <u>***contact***-based</u> approaches.

*: Sample images are from a synthetic fingers from NIST SD300

Tools and Targets

Tools: NIST Fingerprint Registration and Comparison Tool (NFRaCT)

Essentially “diff” for two fingerprint samples

Two biometric samples are loaded in

The samples are registered, and then the software will “compare” them

Sample images are from a synthetic fingerprint generator.

Standardizing Contactless Capture: Publications

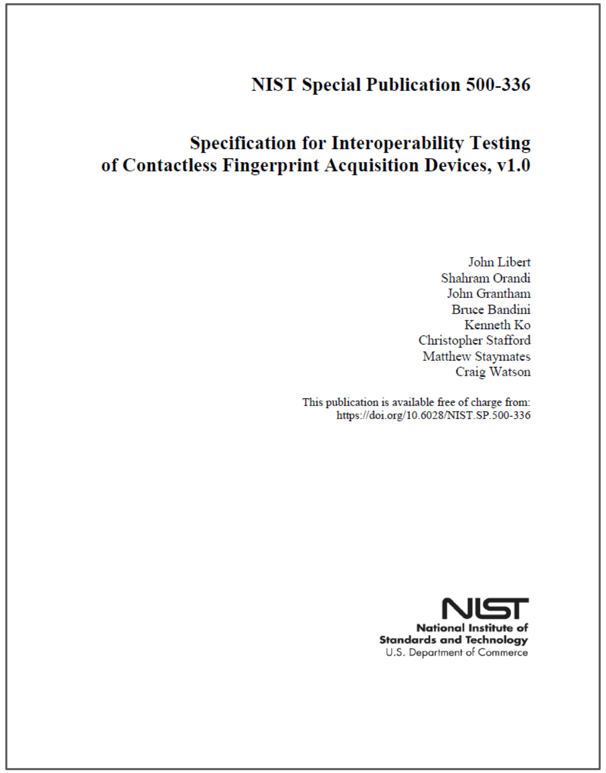
Two special publications.

- NIST SP 500-336: What are the measurands
- NIST SP 500-339: What to measure & what it means

These are in use by 22 partners.

Not a certification by itself – NIST does **not** certify devices..

Initial guidance is for <u>**search**</u> only.



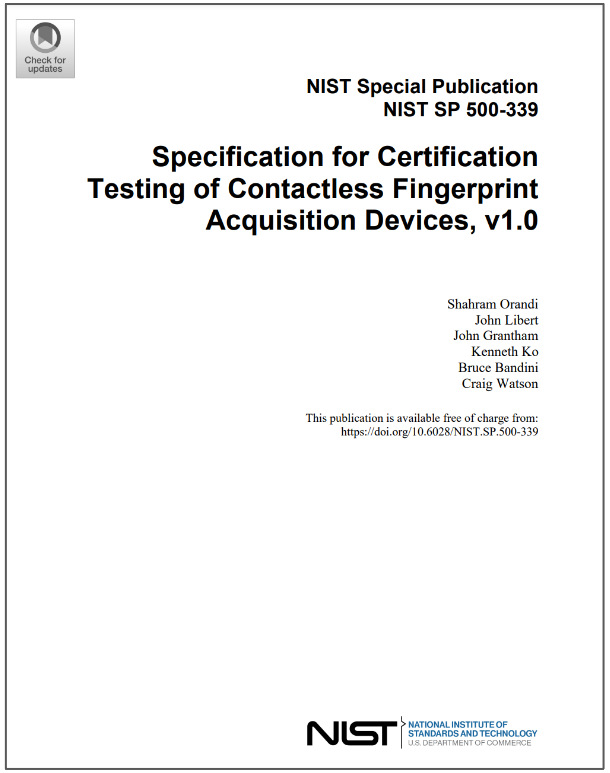
NIST Special Publication 500-336

Specification for Interoperability Testing of Contactless Fingerprint Acquisition Devices, v1.0

John Libert
Shahram Orandi
John Grantham
Bruce Bandini
Kenneth Ko
Christopher Stafford
Matthew Staymates
Craig Watson

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.500-336

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



Check for updates

NIST Special Publication
NIST SP 500-339

Specification for Certification Testing of Contactless Fingerprint Acquisition Devices, v1.0

Shahram Orandi
John Libert
John Grantham
Kenneth Ko
Bruce Bandini
Craig Watson

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.500-339

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

Contactless Fingerprint Data Interchange

NIST Special Publication 500-334

In 2020 NIST began working with 68 partners spanning 30 organizations (national and international) including various government agencies, to develop this guidance.

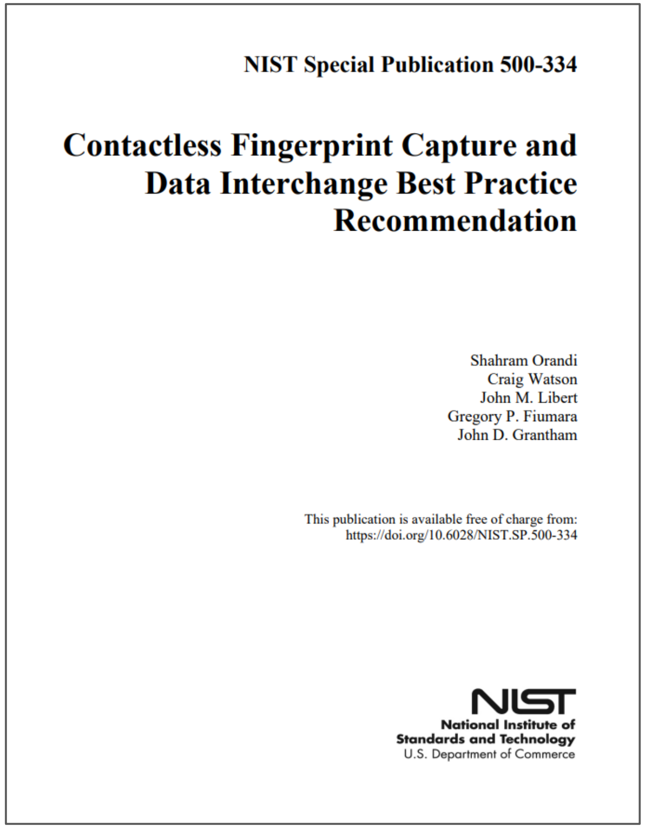
Published in March 2021 (https://doi.org/10.6028/NIST.SP.500-334).

Allows for ***consistent*** data with interchange, and traceability.

Highlights include:

-Informative (for now)

- -New impression types (deprecated 4, added 2) [Isolation]
- -Changes to Make Model Serial number [Traceability]
- -Provision for Raw Sensor Data in Type-20 record [“insurance policy”]
- -Can be normative later


NIST Special Publication 500-334

Contactless Fingerprint Capture and Data Interchange Best Practice Recommendation

Shahram Orandi
Craig Watson
John M. Libert
Gregory P. Fiumara
John D. Grantham

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.500-334

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

Summary Conclusion

- Worked with CRADA partners to understand the technology and development measurements.
- We now have something new with added flexibility and utility (possibly beyond fingerprints):
  - Special Publication #1: Contactless measurands defined, NFRaCT usage how-to (SP500-336)
  - Special Publication #2: Defines pass/fail criteria and required test steps. (SP500-339)
- Data interchange updates – SP500-334
- Targets are still being developed
- We are now in a phase of supporting research partners in testing their devices and supporting partner agencies and stakeholders in pilot testing the specification.
- We are always looking for more test partners on this and welcome all.

For info: please contact us at
fastcap@nist.gov

Human Interactions and Biometrics: Usability

What is Usability?

Usability:
the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use**.**
ISO 9241-11:1998.

Must Understand:

- **Users** – Travelers, operators, examiners, users with disabilities
- **Context** -- Environment, motivation, cognitive load
- **Tasks** -- Acquisition/capture, training, tools
- **Usability metrics** – throughput, accuracy, satisfaction

Why Champion the Human in Biometrics?

January 4, 2004: US began collecting fingerprints and a digital photo of all entering foreign travelers

But the biometrics community forgot about the user

The Result:

(National Science and Technology Council [NSTC], 2008)

Long lines

Confusion of travelers

Overall distrust of the system

What differentiates usability testing from performance testing ?

1. Observation
2. Listening
3. Measuring properties of affordance
4. Interaction of user and device
5. Emphasis that users are not wrong
6. Performance measures are not the whole story

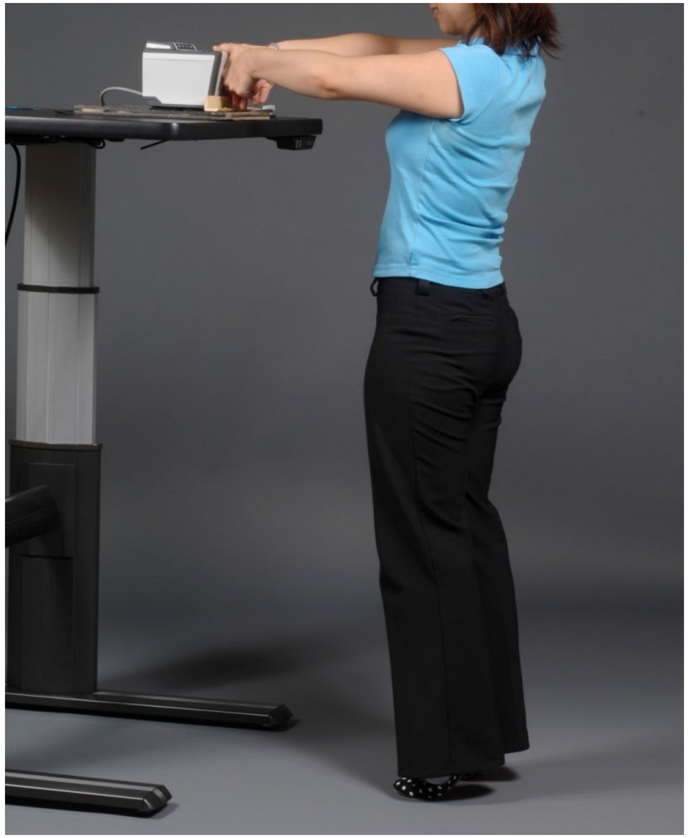
Know Your User

- Sit in an airport and watch
- Performa sufficient testing
- Observe users in action
- What do and don’t they do?

Source: https://www.cbp.gov/newsroom/local-media-release/cbp-introduces-simplified-arrival-denver-international-airport

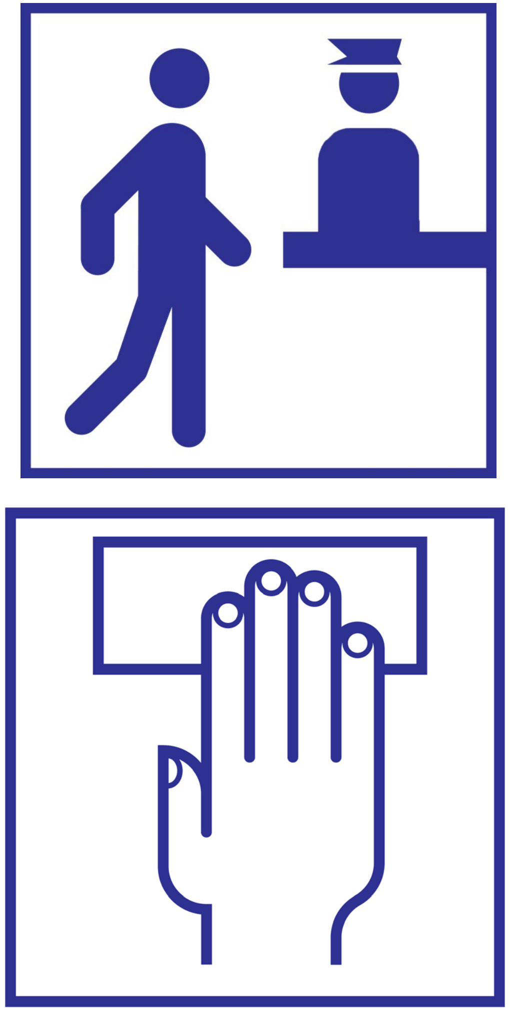
Observing Users is the Key

Taller participants struggle with short counters and scanner angle.

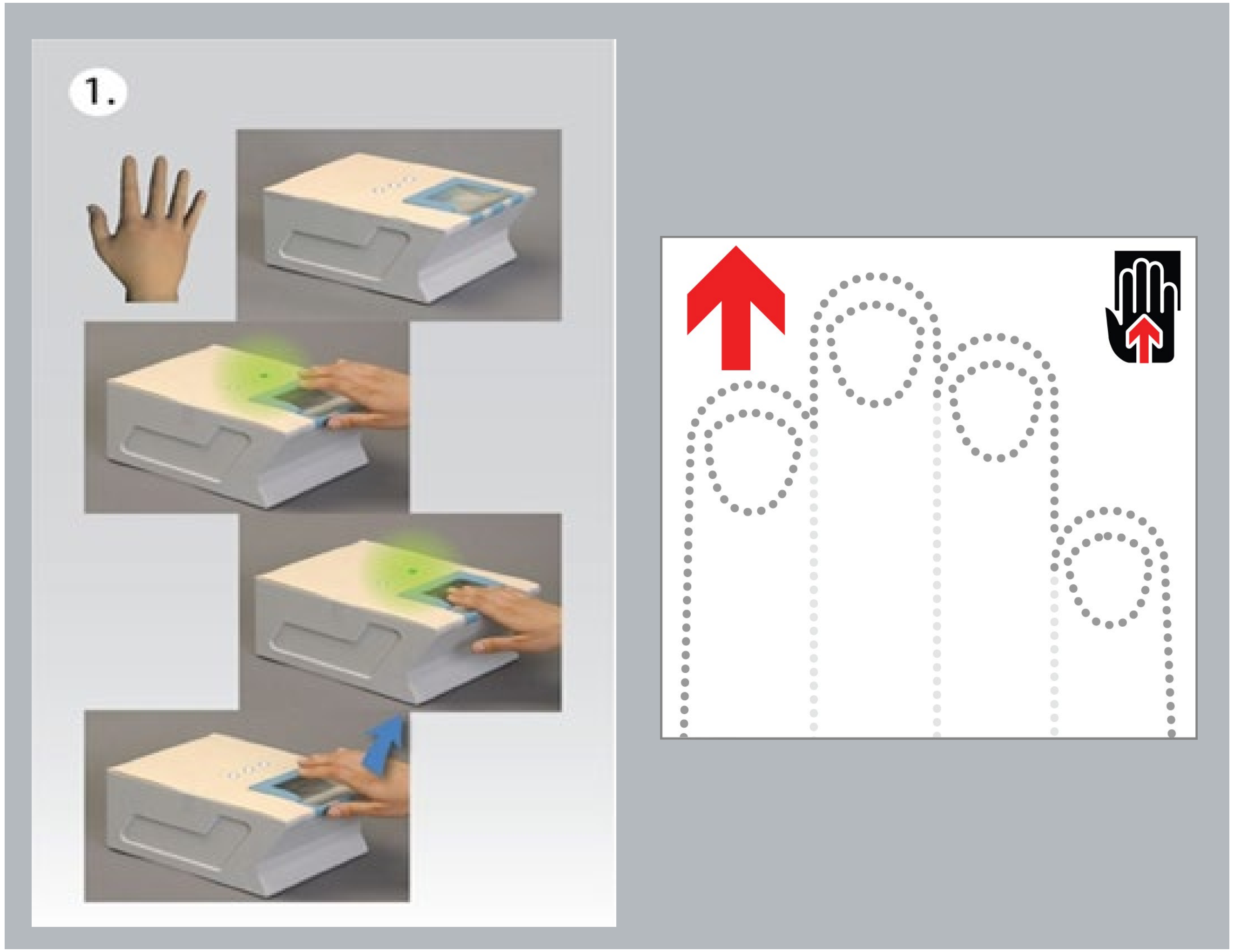


Shorter participants struggle with tall counters and flat scanners.

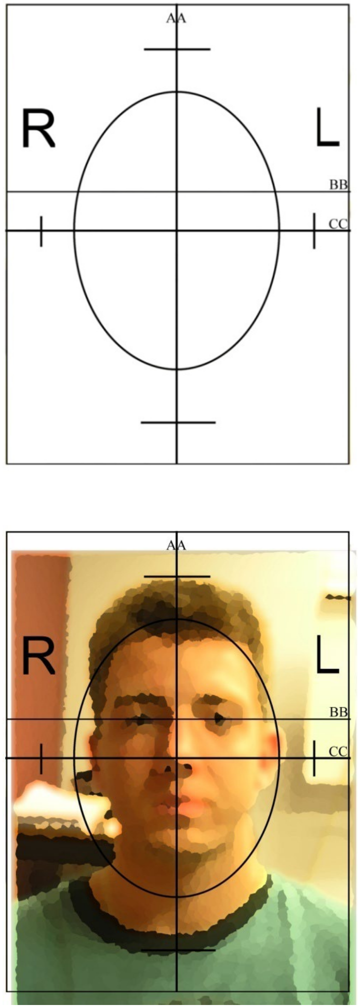
Designed Instructions and Feedback



Symbols



Instructions



Templates

Human Interaction with Device

Past experiences influence use

- **Mental Models** :
  Fingerprint Collection involves pressing fingers against a surface

- **Attitudes**:
  Process is “daunting”, “feel like criminal”, takes a long time, requires many re-tries

- **Behaviors:**
  - Simple instructions – place yellow feet on the floor to indicate where to stand
  - Use graphics/symbols as instructions – ISO 24779 series
  - Provide feedback – when to start process, next step, when are you are finished

Participants placed their hands on the glass surface of the contactless scanners.

Biometric performance is impacted by

User Characteristics

- Age
- Gender
- Height
- Experience (Trust)
- Ability
- Perception

Biometric System Factors

- Ergonomics
- Affordance
- Instructions and Feedback
- Accessibility

Championing the Human in Biometrics has demonstrated:

Human Factors affect biometric performance

- Time required to collect the image
- Quality of the collected image

Which in turn affect system performance

- Throughput
- Matching
- Cost

Q&A (5 minutes)

Factors limiting face recognition

NIST

1. Image quality

2. Twins

→ False Positives

Source: Notre Dame's Twins Day Collection

3. Long-run Ageing

Similar

Less Similar

2002-08 2004-10 2010-05 2012 2013-08 2018-06 2020-04

→ False Negatives

4. Demographics (False Negatives Differentials)

5. Demographics (False Positives Differentials)

6. Human Ability

7. Presentation Attack

8. Morphing

Morph of US Presidents 43+44

NIST International Face Performance Conference 2025

When: April 1 – 3, 2025
Where: In-person @ NIST + remotely over Zoom

Potential Topics:

- Quality Assessment
- Law Enforcement Best Practices
- EU Regulations
- Limits of performance
- Demographics
- Morphing
- Presentation Attack Detection
- Age Estimation and Verification
- Others…

IFPC 2022 website

REGISTRATION IS FREE AND WILL OPEN SOON

Homeland Security Science and Technology

NIST INFORMATION TECHNOLOGY LABORATORY

European Association for Biometrics eab

THANKS

CRAIG.WATSON@NIST.GOV

PATRICK.GROTHER@NIST.GOV

MEI@NIST.GOV

CRAIG WATSON

PATRICK GROTHER

MEI NGAN

https://pages.nist.gov/biometrics-edu/presentations/id4africa_nist_biometrics.pdf