MEASUREMENT OF PERFORMANCE IN BIOMETRIC EXIT

PATRICK GROTHER
NIST

IFPC
OCTOBER 27, 2020

Overview

- FRVT
  - Gains in face recognition
- Biometric EXIT in the USA
- Performance questions
- Performance testing standards

- Simulation with actual image data
  - The algorithm matters
  - The population matters

ONGOING BENCHMARKS

1. FRVT 1:1
Core Biometric Operation

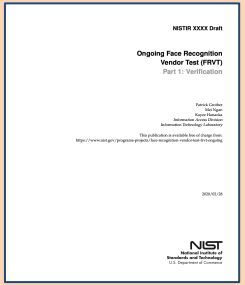
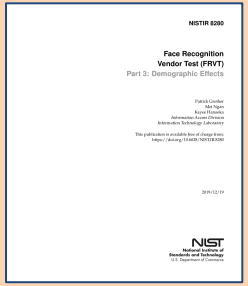
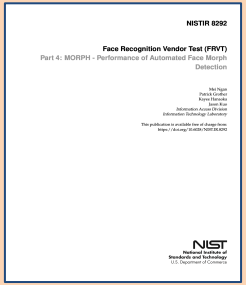
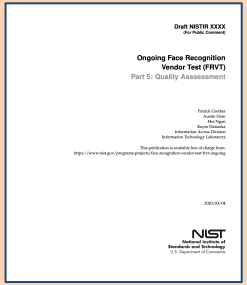
2. FRVT 1:N
Search Performance

3. FRVT Morph
Morphed Photo Detection

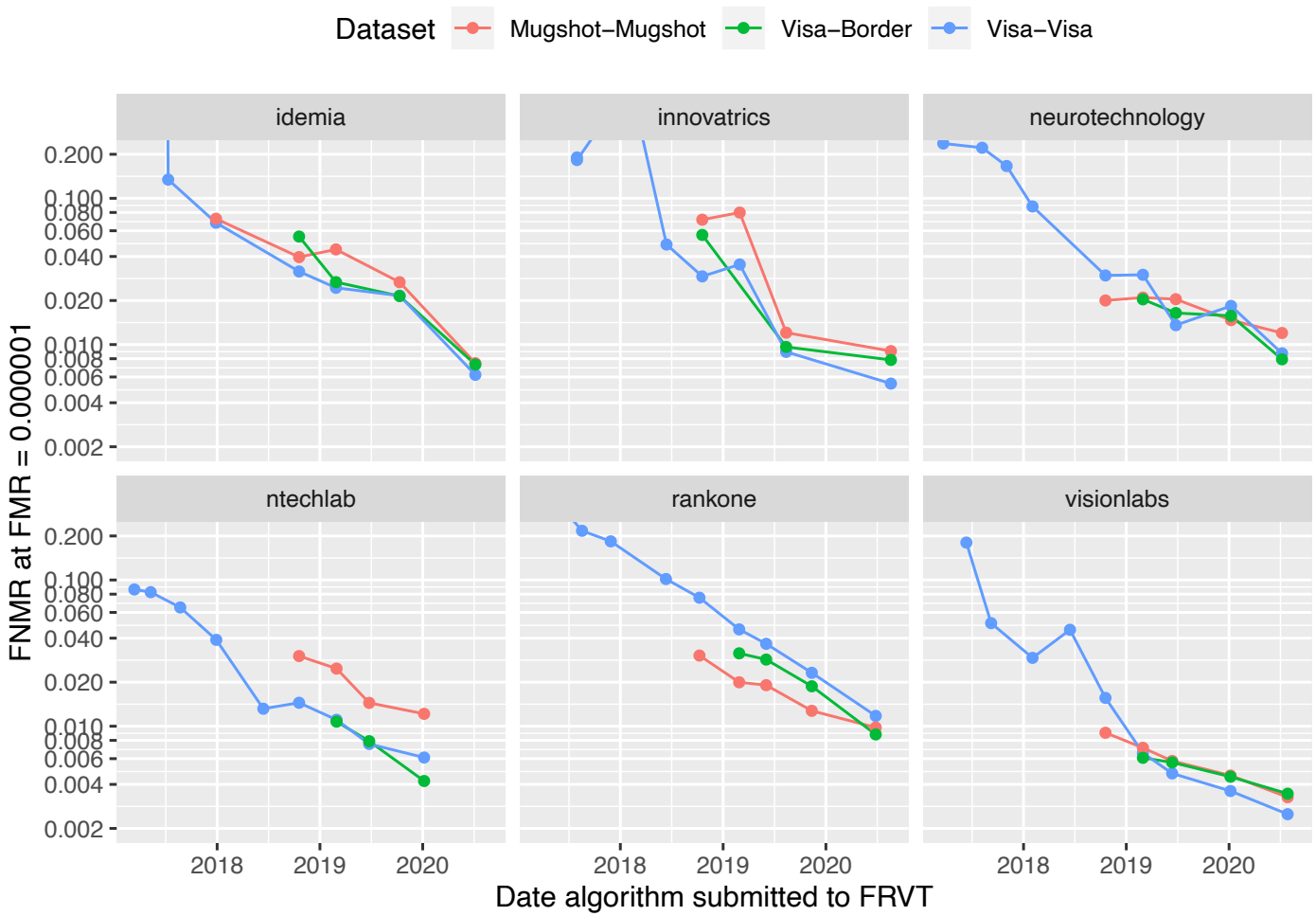
4. FRVT Quality
Automated Quality Assessment

FRVT
Face Recognition Vendor Test

CURRENT PRODUCTS

Part 1: Performance of 1:1 Verification Algorithms	Part 2: Performance of 1:N Identification Algorithms	Part 3: Demographic Effects in Face Recognition	Part 4: Performance of Morph Detection Algorithms	Part 5: Performance of Image Quality Assessment Algorithms	Part 6: Performance of Face Recognition with Face Masks	Part 7: Performance of Face Recognition on Twins
						
Last: 2020-10-09 Next: 2020-11-15	Last: 2020-03-27 Next: 2020-11-05	Last: 2019-12-19 Next: 2021-02 est.	Last: 2020-07-24 Next: 2020-11 est.	Last: 2020-07-27 Next: 2020-11 est.	Last: 2020-09-18 Next: 2020-11 est.	Last: Next: TBD

Industry-wide gains in accuracy



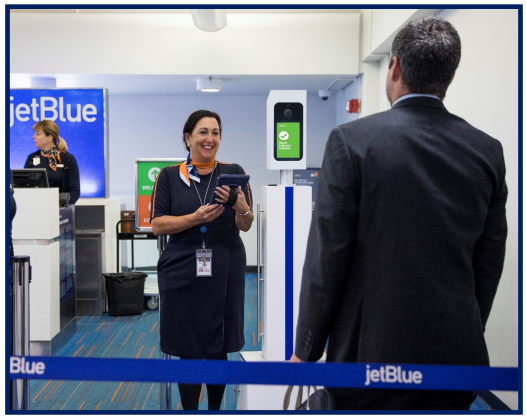
- FNMR at low FMR
- Gains over time ~ 10% ⟶ below 1%
  - Quality, ageing, children
- Capability varies by developer
- FNMR < 0.004
- Visa-visa and mugshot-mugshot accuracy comparable to visa-border.

New FRVT 1:N Benchmarks

FNIR @
FPIR 1/333

	Mugshot 12M		Mugshot Webcam	Mugshot Profile	Visa-Border	Visa-Kiosk	Mugshot Ageing
Algorithm	Mugshot Mugshot N = 12000000	Mugshot Mugshot N = 1600000	Mugshot Webcam N = 1600000	Mugshot Profile N = 1600000	Visa Border N = 1600000	Visa Kiosk N = 1600000	Mugshot Mugshot 12+YRS N = 3000000
xforwardai 000	0.0217(14)	0.0099(15)	0.0437(20)	0.1228(3)	0.0151(11)	0.1496(1)	0.0986(14)
imperial 000	0.0340(31)	0.0151(28)	0.0525(23)	0.9923(106)	0.0298(25)	0.2003(2)	0.1565(23)
rankone 009	0.0258(23)	0.0124(22)	0.0597(29)	0.8180(30)	0.0427(29)	0.2463(3)	0.1209(19)
kedacom 001	0.0293(25)	0.0167(31)	0.0621(31)	0.9832(85)	0.0476(33)	0.2786(4)	0.1038(15)
dahua 002	0.0218(16)	0.0099(14)	0.0373(14)	0.3540(12)	0.0127(9)	0.2844(5)	0.0962(12)
visionlabs 009	0.0074(6)	0.0030(7)	0.0210(7)	0.1977(7)	0.0064(3)	0.2878(6)	0.0333(6)
lookman 005	0.0390(38)	0.0226(40)	0.0740(41)	0.9809(81)	0.0541(38)	0.2890(7)	_
visionlabs 008	0.0194(12)	0.0086(11)	0.0416(16)	0.1799(5)	0.0121(8)	0.3200(8)	0.0821(10)
lookman 3	0.0555(47)	0.0329(62)	0.0961(69)	_	0.0722(43)	0.3288(9)	0.1902(35)
pixelall 003	0.0314(30)	0.0140(24)	0.0562(28)	0.9974(126)	0.0287(23)	0.3858(10)	_
ntechlab 007	0.0360(33)	0.0175(36)	0.0534(25)	0.4694(15)	0.0245(18)	0.4499(11)	0.1824(30)
ntechlab 008	0.0218(15)	0.0099(16)	0.0364(13)	0.1998(8)	0.0284(22)	0.4507(12)	0.1718(27)
paravision 005	0.0065(5)	0.0030(6)	0.0199(6)	0.2335(9)	0.0098(7)	0.4917(13)	0.0336(7)
ntechlab 6	0.0528(43)	0.0267(45)	0.0744(42)	0.4382(13)	0.0456(30)	0.4942(14)	0.2945(56)

Immigration Exit



- Cameras: Multiple developer implementations
- Population:
  - Those expected on flight per airline-provided flight manifest.
  - US citizens included
- Biometric:
  - Face images from all known prior encounters of subject
  - Hundreds of people
  - Thousands of images
  - For USCs, prior passport images.
- Search live face against database
  - Zero or more attempts per passenger
- Cloud-based FR
  - An NEC algorithm
  - Image sent over high-speed low latency network. Round-trip transaction ~ 1 second.

- Airline, airport, government partnership
- Application
  - Facilitation of traveler's recording their exit
  - Physical access control

- Errors and resolution
  1. False negative:
     - Failure to record biometric exit, revert to biographic?
     - Revert to traditional paper-based boarding process
  2. False positive from legitimate passenger:
     - Traveler boards plane; may be detected if and when actual traveler subsequently boards vessel.
     - False negative for the legitimate passenger (case 1)
  3. False positive from non-passenger:
     - Traveler boards plane; May be detected if and when actual traveler subsequently boards vessel.
  4. Active attack success
     - Incorrect recording of biometric exit

1:N for “paperless” single-factor authentication

JetBlue –
Paperless Boarding at BOS

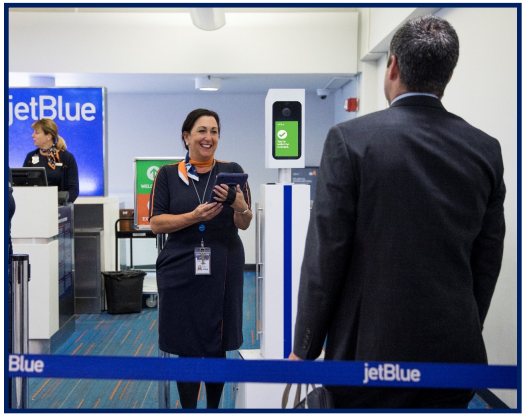


Image from CBP presentation
“Airlines integrate facial biometrics and use CBP’s agnostic matching service to board aircraft.”

Los Angeles – E-gates

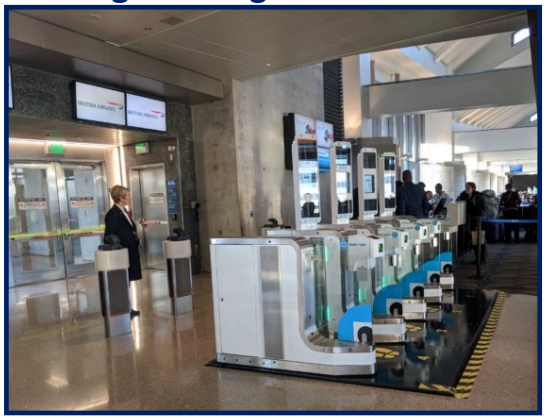


Image from CBP presentation

Image from MWAA website
https://www.mwaa.com/about/veriscan-quickest-path-biometric-compliance

“veriScan eliminates the need for travelers to present their passport and boarding pass at the gate. **The passenger’s face serves as both of these required documents**, resulting in a hands-free, touchless boarding process.”

Single factor authentication:

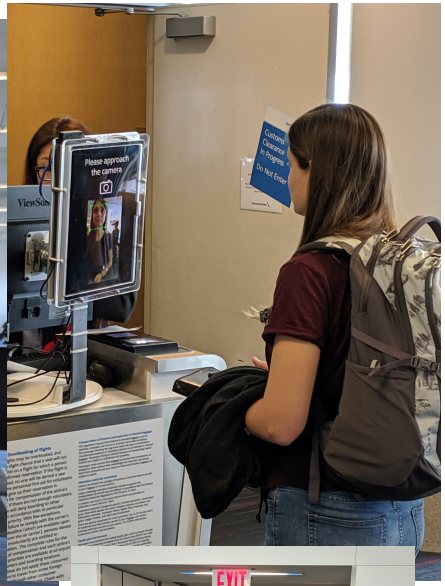
1. Something you are: Successful face identification of live-image to database

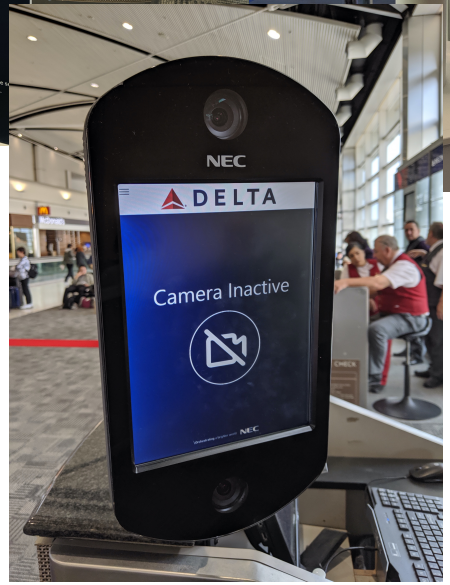
Possibilities for future paperless travel

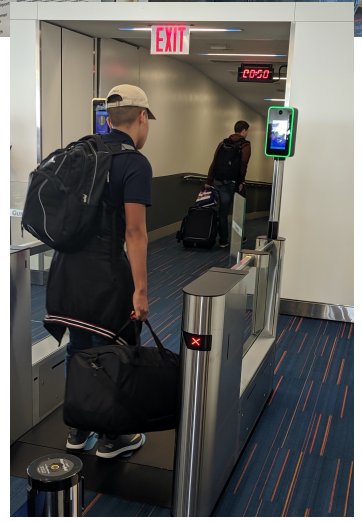
#	Which border	Step	1:1 Verification or 1:N Identification	Where	Enrolled database	FNIR, FPIR Requirements
1	AIR	Initial verification against document	1:1 against passport or driving license	Check-in Automated bag drop	N = 1	?
2	AIR	Is passenger allowed airside?	1:N	TSA Screening checkpoint	N ~ 10^5	?
3	AIR	Duty free shopping	1:N	Air-side shops	N ~ 10^5	?
4	AIR	Lounge access	1:N	Airline lounges	N ~ 10^4	?
5	AIR	Record immigration exit	1:N	At boarding gate	N < 500	FNIR < 0.03[+] ?

[+] 2007 Visa Waiver Modernization Act

1:N Face:
Diverse hardware, common matcher

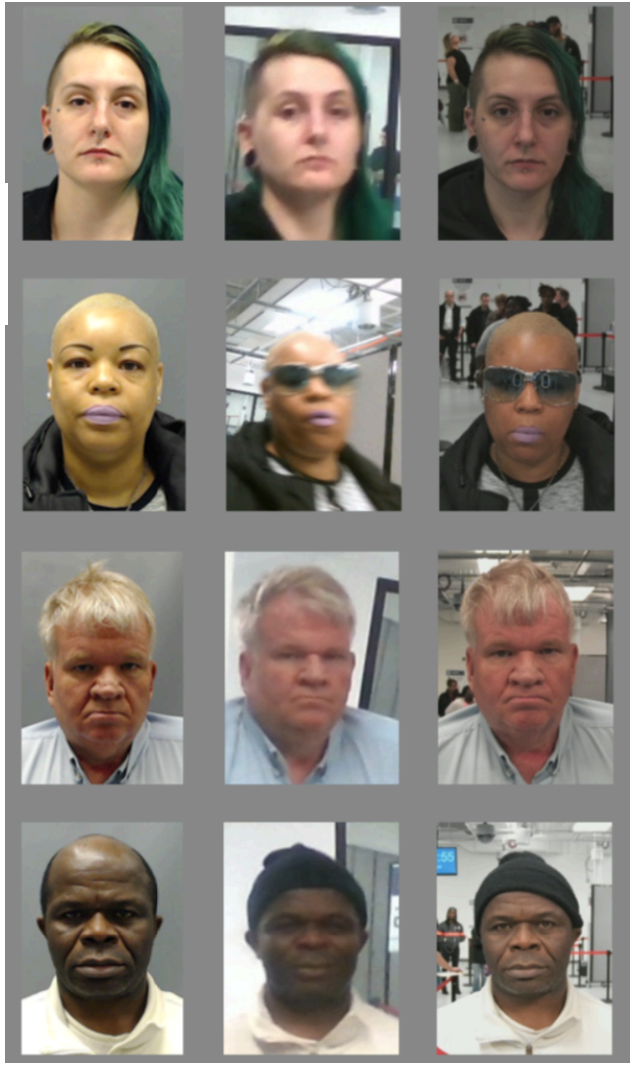






The camera matters

Fig. 8. Example imagery from enrollment, System 03, and System 07. Subjects shown are (in order from top), self-identified as White with highest skin reflectance, self-identified as Black or African-American with highest skin reflectance, self-identified as White with lowest skin reflectance, and self-identified as Black or African-American with lowest skin reflectance.



Demographic Effects in Facial Recognition and their Dependence on Image Acquisition: An Evaluation of Eleven Commercial Systems. Cynthia M. Cook, John J. Howard, Member, IEEE, Yevgeniy B. Sirotin, Member, IEEE, Jerry L. Tipton, and Arun R. Vemury. IEEE Transactions on Biometrics, Behavior, and Identity Science, February 2019

How well does it work? What does that mean?

- How accurate is it?
- How fast is it?
- How secure is it?

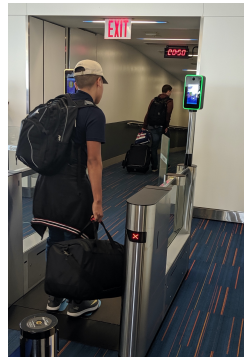
- Is it more accurate than biographic exit?
- Is it trustworthy?
- What is the visa-overstay rate?

- Is everyone enrolled, enrollable?
- How many opt out at exit?
- How many are steered out by airline staff?
- How many authenticate on a second or third attempt?
- Does it depend on
  - Which camera?
  - What time of day?
  - The population?
    - Age, sex, race, height?
    - Beards, sunglasses, face masks?

What kinds of tests might you run?

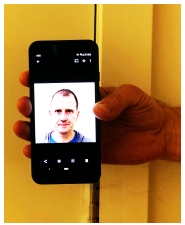
- Technology tests
  - Often algorithm running on offline images
    - Example: NIST FRVT
  - With some metadata on population, location, equipment etc.
  - Pros: Repeatability, scalability, significance, low cost
  - Cons: No info. on capture, inc. failure, transactional aspects, process speed
- Scenario tests
  - Volunteer human-in-the-loop tests in controlled env. with capture devices and real-time matching
    - Example: DHS S&T MdTF
  - Pros: Capture user-interaction, feedback, speed, satisfaction; can compare cameras, processes Cons: Expense, scalability to large populations, not readily repeatable

- Operational tests
  - In-the-field, actual population and system, with instrumentation
  - Pros: Realism; online results + offline possibilities
  - Cons: Undetected impostors, instrumentation ⟶ Heisenbugs, lack of control, repeatability
- Security tests
  - Circumvention
  - Presentation Attack Detection



MASK

REPLAY



Standards:

Biometrics Performance Testing and Reporting

- ISO/IEC 19795-1
  - Principles and Framework
- ISO/IEC 19795-2
  - Technology and Scenario Testing
- ISO/IEC 19795-6
  - Operational Testing
- ISO/IEC 30107-3
  - Presentation Attack Detection Testing
- Plus many more:
  - Environmental, Interoperability, PACS, On-Card Comparison, Template Protection, Operator-led, Mobile, Demographics
- http://webstore.ansi.org | local standard bureaus.

Standards developed in ISO/IEC JTC 1 SC 37
Participation: Email patrick.grother@nist.gov

Simulated EXIT experiment

Enrol

- Generate 498 galleries, one per departing flight
  - N = 480 people in each gallery
  - One immigration ENTRY photo per person
  - All travelers have same region of birth in their travel document.
  - No attempt to control age, sex.
- Total num images: 825 977
- Total num people: 81 114

1:N Search

- One search set
  - K = 127 258 EXIT Images of 123 074 people
  - Num. mated searches per flight ~ 480
  - Num. non-mated searches ~ K – 480 = 126 778
- Num images by region

EUROPE	52559	N. AMERICA	14104	E. ASIA	17882
N. AFRICA	464	CARIBBEAN	2714	OCEANIA	8717
SUB-SAH. AFRICA	1893	C. AMERICA	8789	S. ASIA	5276
MID EAST	3487	S. AMERICA	11373		

Entry-Exit Test Data

ENTRY

REFERENCE SAMPLES (Examples are very similar, not actual, taken from NIST Special Database 32)

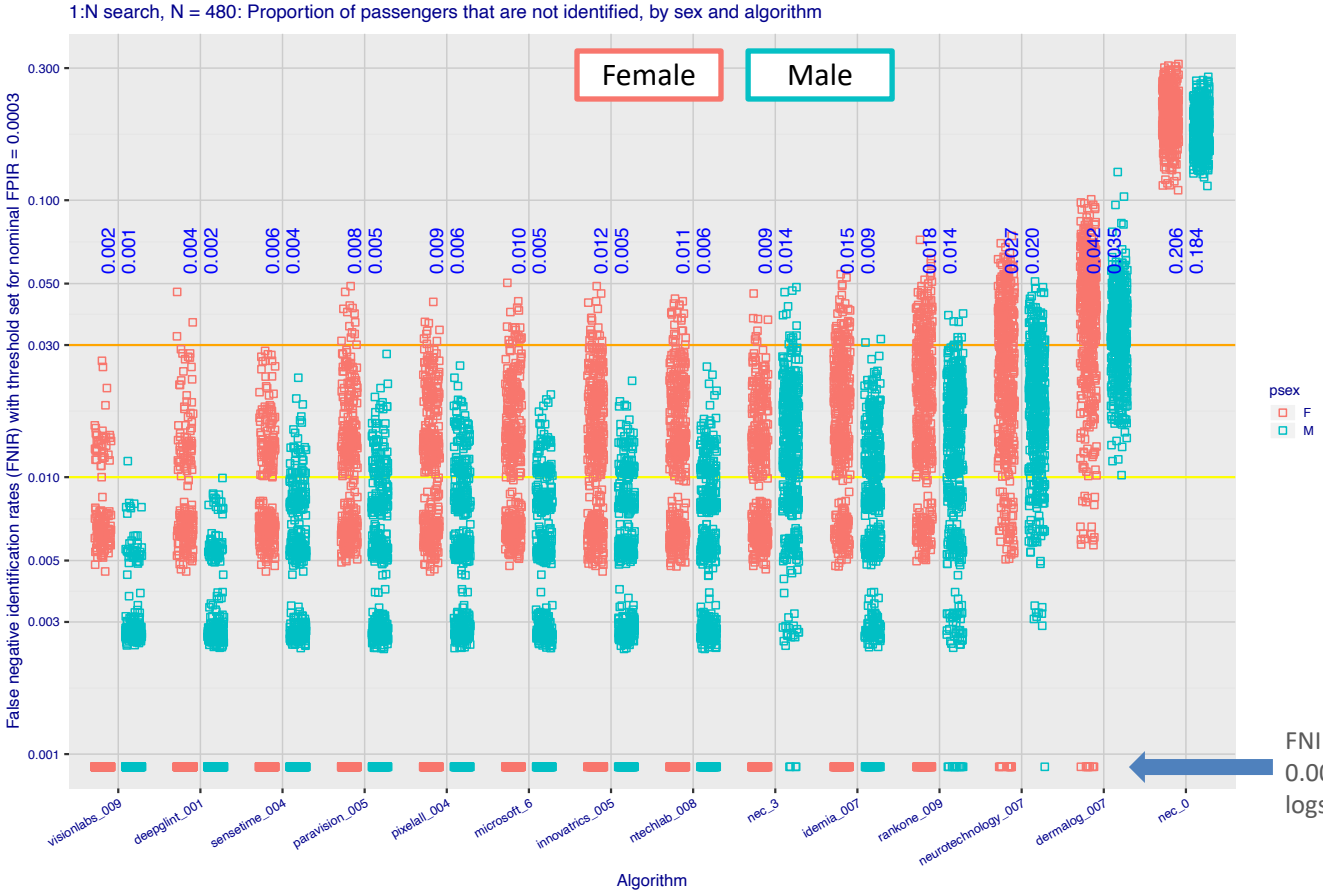
EXIT

VERIFICATION SAMPLES

- Better than entry w. r. t
  - Illumination
  - Compression
  - Pose
  - Size uniformity and margins.
- Cluttered background, some faces
- Pose variation > ISO limits
- Several cameras evident

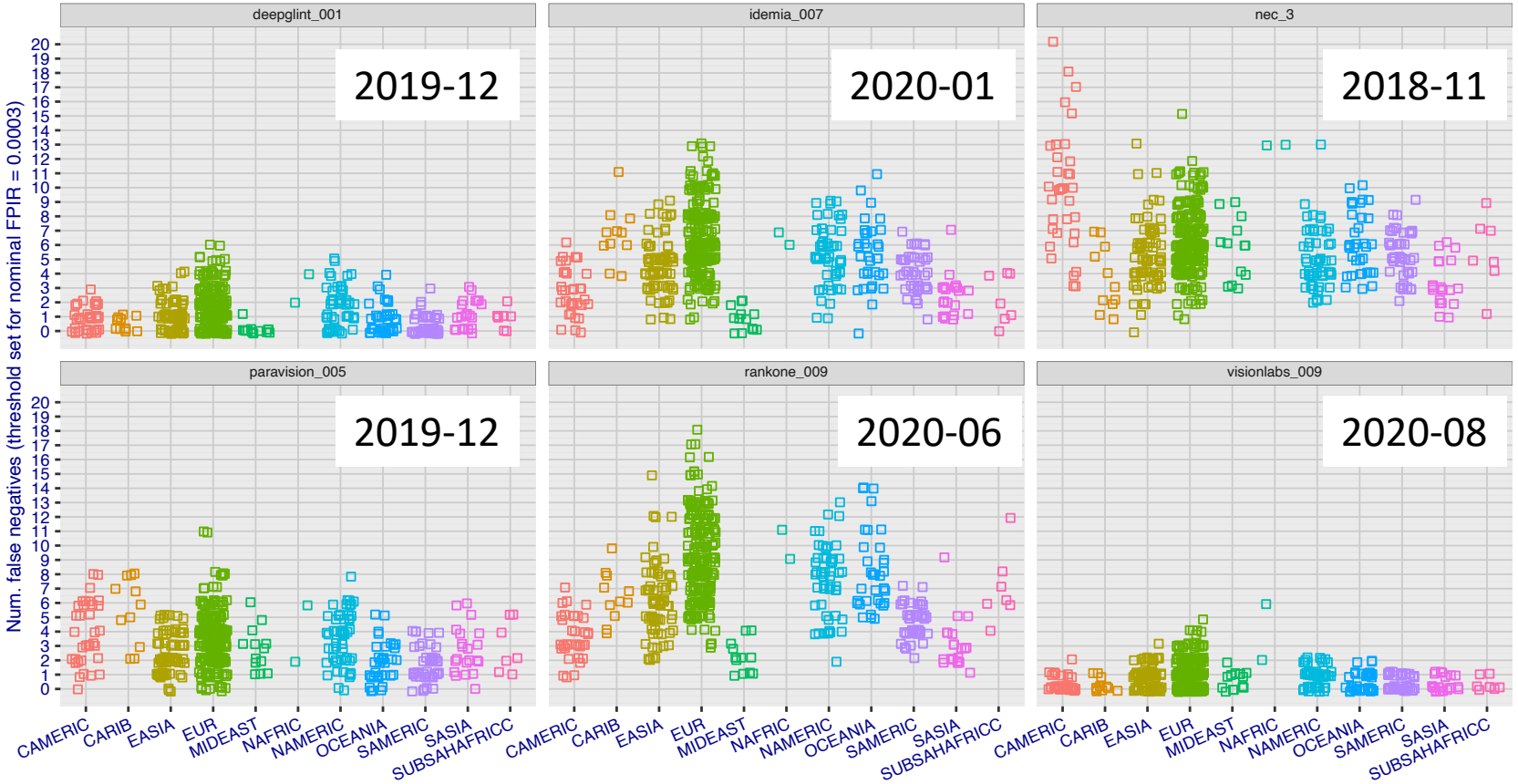
FNIR ~ Proportion of passengers not identified

1. Distribution over flights of FNIR
2. Visionlabs low FNIR
3. FNIR < 3%
4. FNIR(F) often higher
5. Version matters



FNIR = 0 Is coded as 0.0009 for display on logscale

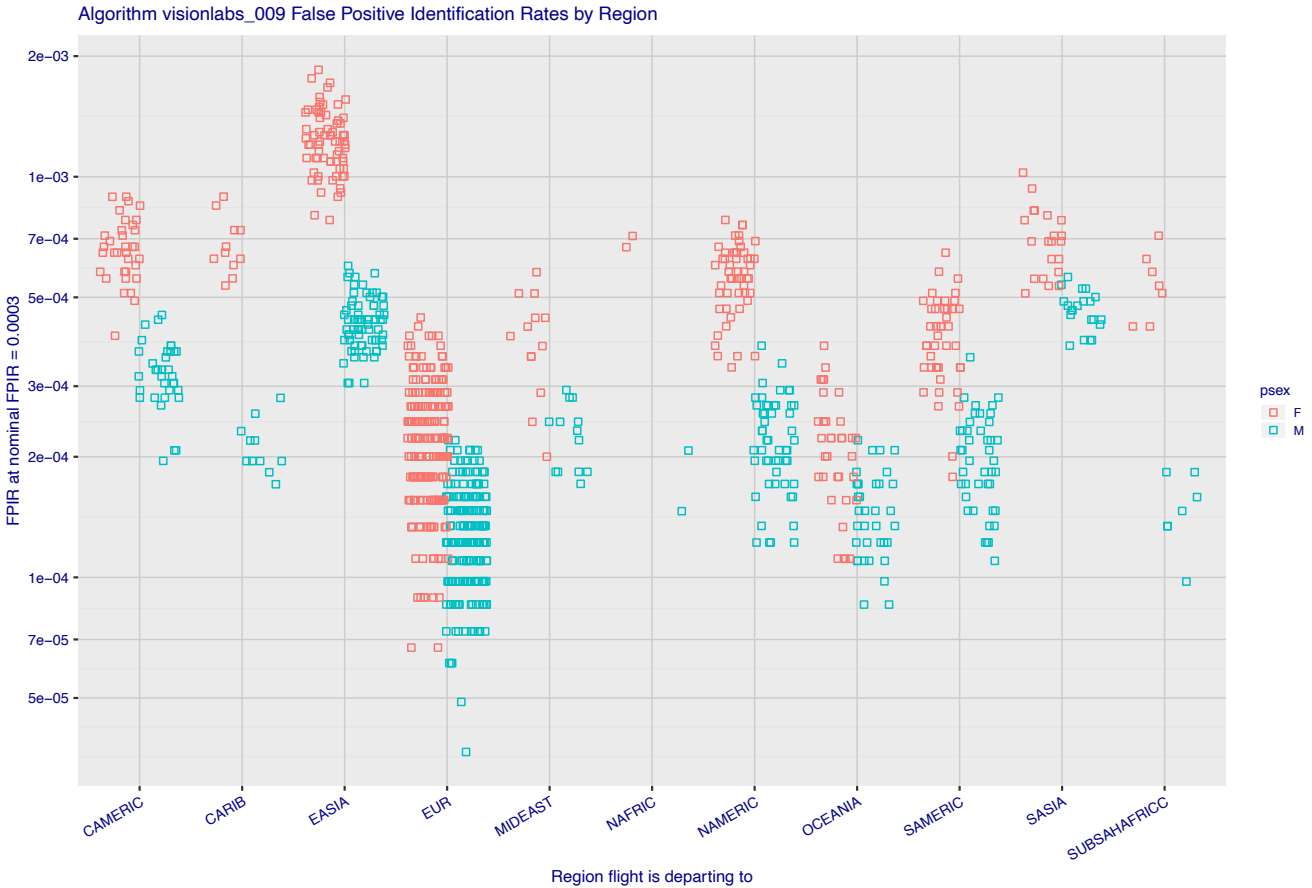
EXIT Search Accuracy by Region and Algorithm



N = 480

False positive identification rates: Visionlabs-009

1. Visionlabs ~ most accurate for FNIR
2. FPIR in East Asian women ~ 10x European Men
3. But FPIR < 1:500 everywhere
4. ΔFPIR is algorithm dependent



Summary

- Gains in face recognition
  - Enables high throughput capture
  - Development rates > Technology refresh rates
- Accuracy can be high, dependent on (in order)
  - The algorithm
  - The population
  - The camera equipment
  - ...
- Tests
  - Offline tests answer some questions economically, and at scale
  - Scenario tests can give estimates of full-system performance
  - Operational tests can answer questions definitively
  - Testing standards exist!

THANKS

PATRICK.GROTHER@NIST.GOV

FRVT@NIST.GOV