FACE STANDARDIZATION

IMPROVING FACE RECOGNITION VIA SPECIFICATION OF IMAGES, MEASUREMENTS ON IMAGES, CAMERAS

PATRICK GROTHER

NIST, U. S. DEPARTMENT OF COMMERCE

IFPC 2020

OCTOBER 28, 2020

Overview

- Problems:
  - Face recognition failures I
  - Face recognition failures II
- Solutions:
  - Documentary standards  ISO/IEC 39794-5 ANNEX D + E
  - Image quality assessment  ISO/IEC 29794-5 FACE IMAGE QUALITY
  - Smarter cameras  ISO/IEC 24358 FACE-AWARE CAPTURE SUBSYSTEM SPECS.
- FRVT Evaluation

MOTIVATION FOR QUALITY
FACE RECOGNITION FAILURES I

NIST Fingerprint Image Quality

NFIQ v1 (2004)

NIST (Tabassi, Wilson)

Quality number {1,2,3,4,5}

Feature extraction: computes appropriate signal or image fidelity characteristics and results in an 11-dimensional feature vector. Features are hand-crafted.

Classifier = Neural network: $R^{11} \longrightarrow I^{1}$ {1,...5} based on various quantiles of the normalized match score distribution.

NFIQ v2 (2016)

NIST + BSI + BKA + Fraunhofer IGD + Secunet + Hochschule Darmstadt

Quality number {0 ... 100}

Feature extraction: computes appropriate signal or image fidelity characteristics and results in an 14-dimensional feature vector. Features are hand-crafted.

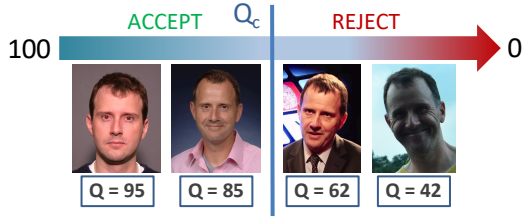
Classifier = Random Forest: $R^{14} \longrightarrow R^{1}$ [0,100] classifies feature vectors two classes with quality value being the probability of class membership. The two classes depend on NFIQ v1 and genuine similarity scores from 9 commercial matchers

ISO/IEC 29794-4:2017 Biometric sample quality -- Part 4: Finger image data

Why measure quality

Scalar quality values (like NFIQ + commercial)

- Image acceptance / rejection decisions during enrollment



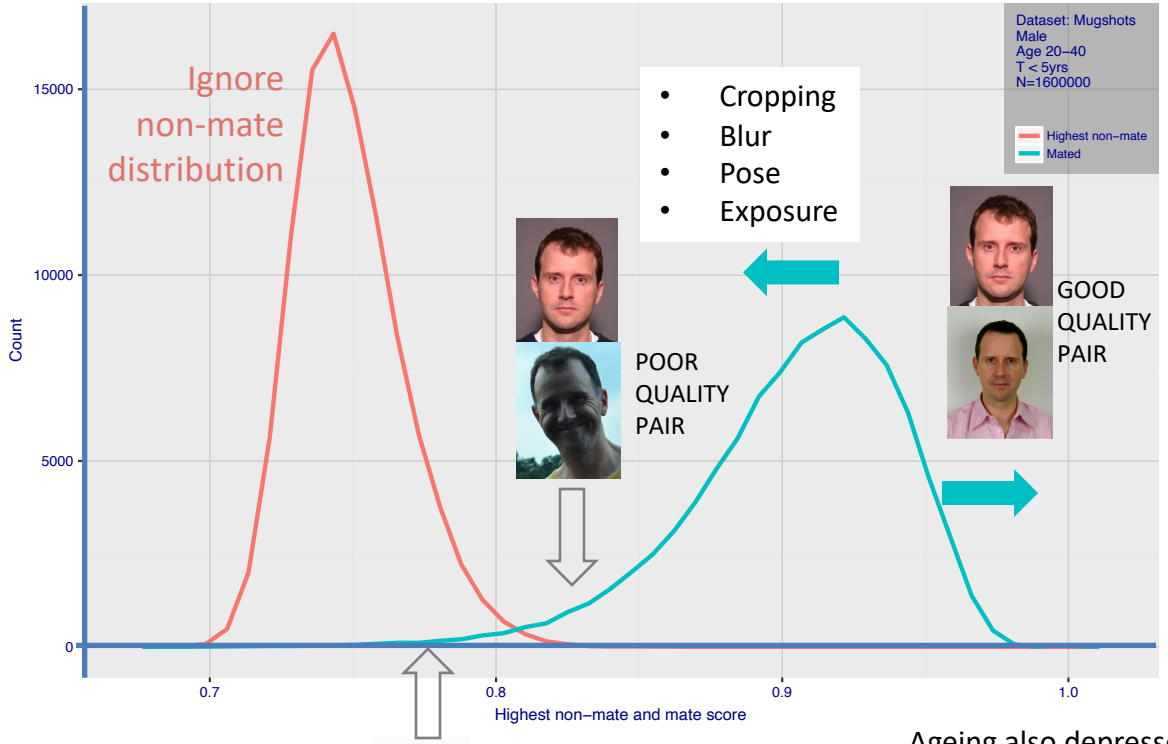
- Sample selection from a capture stream
- As a management indicator
  - Monitor a statistic over time, place, camera, organization etc.
- To augment a multi-biometric fusion process
- To augment human review
- NOT as a replacement for matching

Vector quality values Image analysis (cf. commercial packages)

- Is the image defective, and how?
  - Actionable feedback to operator / subject
- Standards conformance
  - Is the image blurred?
  - Is the person facing the camera?
- Expose method for remediation
  - Is the image defect related to subject mis-presentation?
  - Is the image defect systematic, occurring even with a perfectly presenting subject?
- As a management indicator

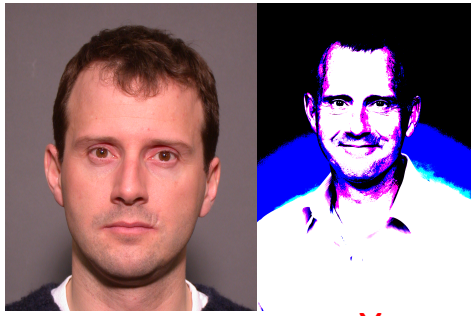
Face Recognition Failures I

False Rejection



Ageing also depresses matching scores independent of quality
⟶ exclude aged pairs from a quality assessment set

Why ignore the impostor distribution?



X

Anecdotally: Some recognition algorithms give false matches from saturated images

- There is an operational role for detecting images likely to give FM

But: These images, both “high quality”, also false match (Why? sisters).

- FM is caused by biology AND quality problems (at least)
- This impedes evaluation

Assumption: It is sufficient to evaluate a QA algorithm on prediction of low genuine scores (from images like X)

STANDARD 1
ISO/IEC 29794-5 FACE IMAGE QUALITY

Face quality standard tests + structure

Unified quality score	6.2
Illumination uniformity	6.3
Illumination uniformity (alt)	6.4
Illumination under-exposure	6.5
Illumination over-exposure	6.6
Illumination over-exposure (alt)	6.7
Illumination modulation	6.8
De-focus	6.9
Image sharpness	6.10
Motion blur	6.11
Edge Density	6.12
Compression	6.13
Unnatural colour and colour balance	6.14
Eyes visible	6.15
Number of faces present	6.16
Inter-eye distance	6.17
Horizontal position of the face	6.18
Vertical position of the face	6.19
Background uniformity	6.20
Pose	6.21
Expression neutrality	6.22
Mouth closed	6.23
Eyes open	6.24
Developer-defined quality score computation	8 and Annex A

» **Description:**

- ISO/IEC 39745-5:2019 in clause D.2.4.2 recommends: The dynamic range of the image should have at least 7 bits of intensity variation (span a range of at least 128 unique values) in the face region of the image. The face region is defined as the region from crown to chin and from the left ear to the right ear.

» **Computation**

- Find and segment the face region
- Recover the image luminance from the encoded data e.g. gamma inversion then Y = 0.2126R + 0.7152G + 0.0722B
- Compute the number of pixels, n_i, whose integer intensity is i
- Compute the number of pixels, $N = \sum_i n_i$
- Compute $p_i = \frac{n_i}{N}$
- Compute entropy $H = -\sum p_i \log_2 p_i$

» **Units of measure**

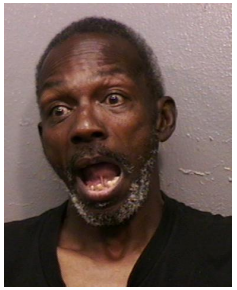
- The unit of entropy are bits.

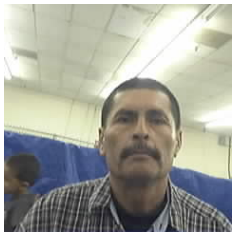
» **Value range and threshold**

- Range: [0,8] for an 8 bit image
- Acceptable: $H \geq 7$, otherwise the image is not correctly exposed.

Face Quality Role #1: Face Image Quality Attributes

Vector Quality: Quantitative checks of subject and image properties

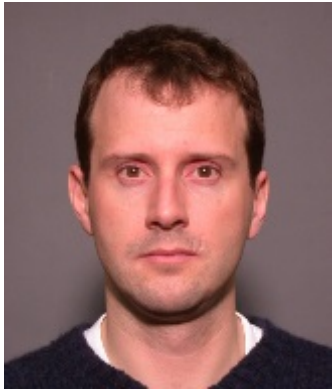




STANDARD 2 -- ISO/IEC 24358
FACE-AWARE CAPTURE SUBSYSTEM SPECIFICATIONS

ISO/IEC 24358

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



Images from presenter

Extant Problems:

a) Non-frontal faces
b) No-faces, multiple-faces
c) Over- and under-exposure
d) Human review errors
e) Demographics
e) Morphing is possible
f) Inadequate for presentation attack detection

Potential Solutions:

a) Face pose detector
b) Face detectors
c) 12 bits or closed-loop exposure control
d) Higher resolution, better compression, 3D
e) Crypto for tamper-proofing
f) Automatic upload to issuing authorities

ISO/IEC 24358

MOTIVATION FOR MORE RESOLUTION
FACE RECOGNITION FAILURES II

Scenario: Identical Twins

Probe is an identical twin

Gallery Size: 1.6 million

Algorithm	Rank of sibling	Score	FPIR
Microsoft	1	0.78	0.0007
NEC	1	0.77	0.0010
Idemia	1	3066	0.0007

Almost all algorithms give high scores

Candidate List

Rank 1

...

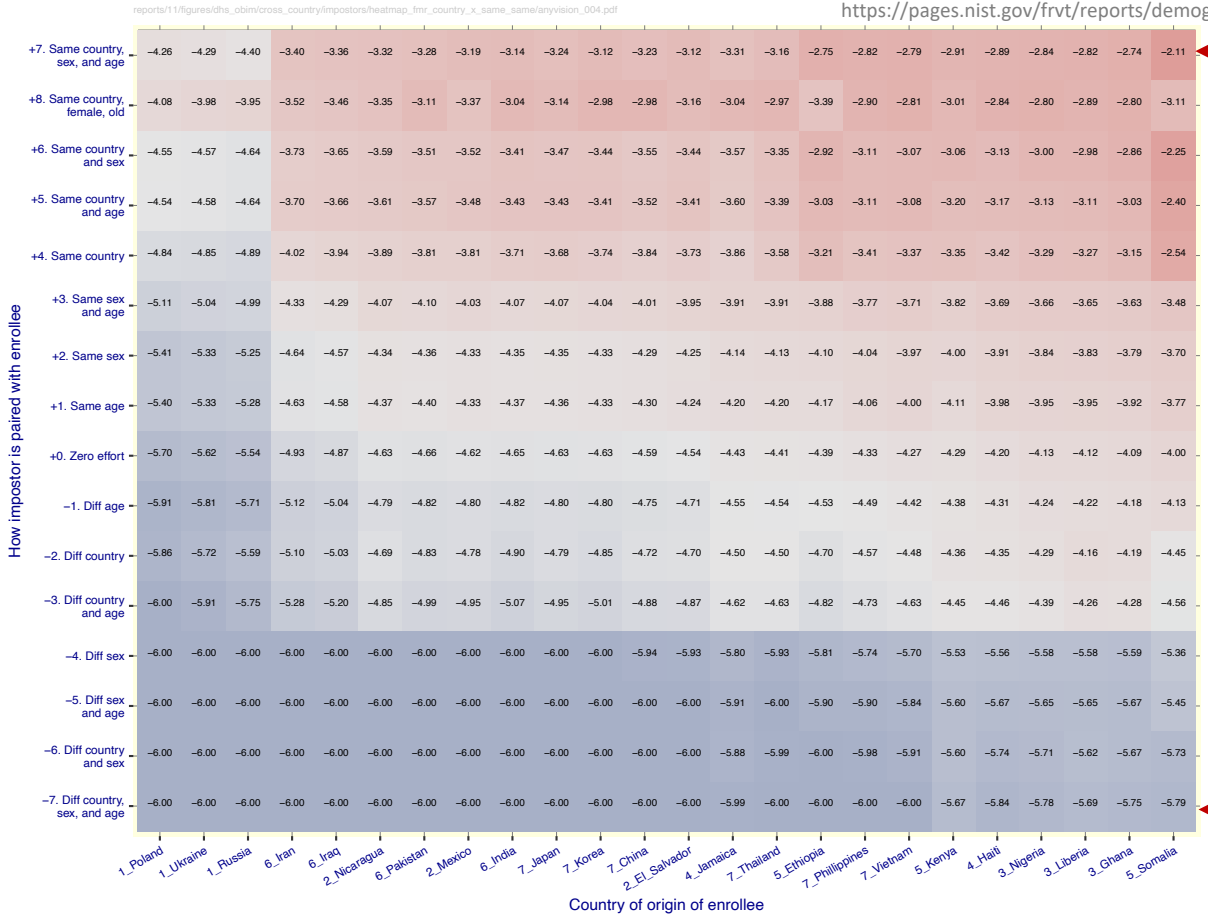
...

”Broad homogeneity”:: FMR increases with sameness of demographics

Source: NISTIR 8280 Annex 8
https://pages.nist.gov/frvt/reports/demographics/annexes/annex_08.pdf

COMPARE PHOTOS OF PEOPLE FROM **SAME** AGE, SEX, RACE

COMPARE PHOTOS OF PEOPLE FROM **DIFFERENT** AGE, SEX, RACE

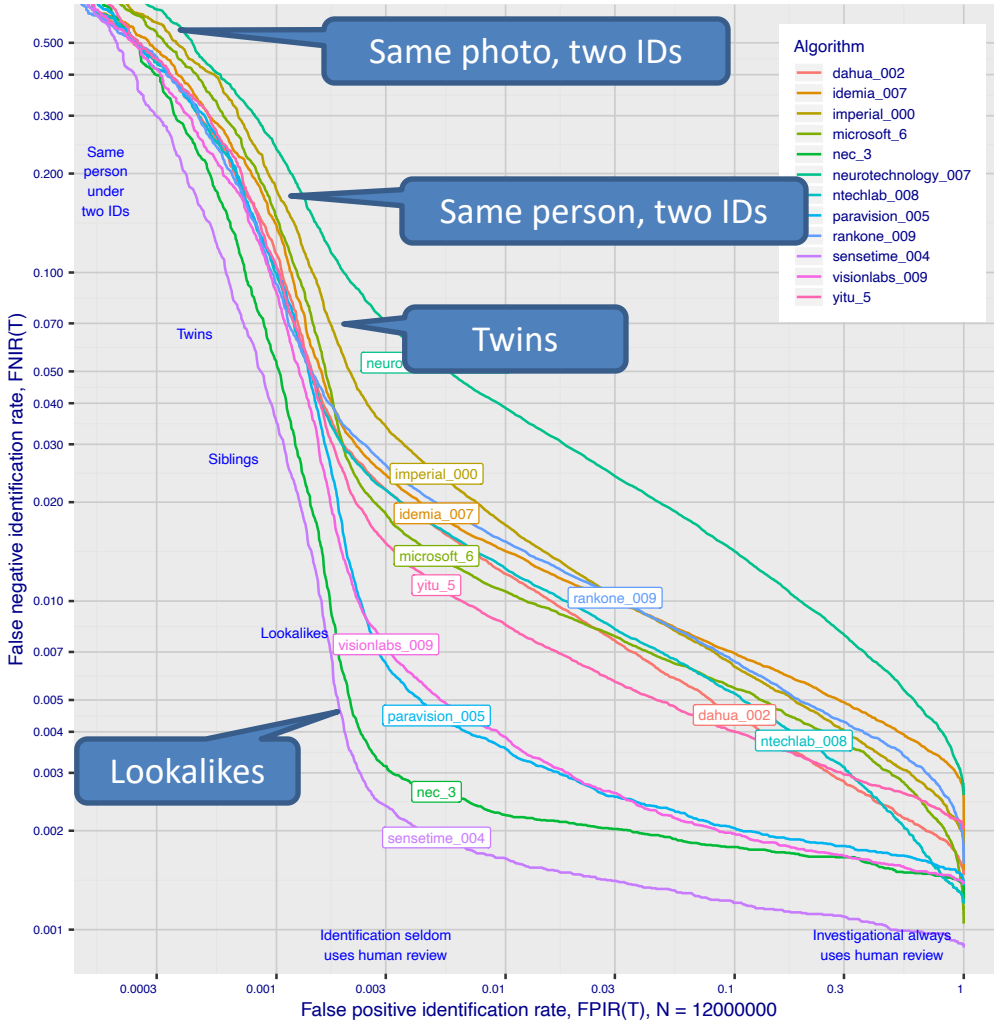


FMR spans two+ orders of magnitude

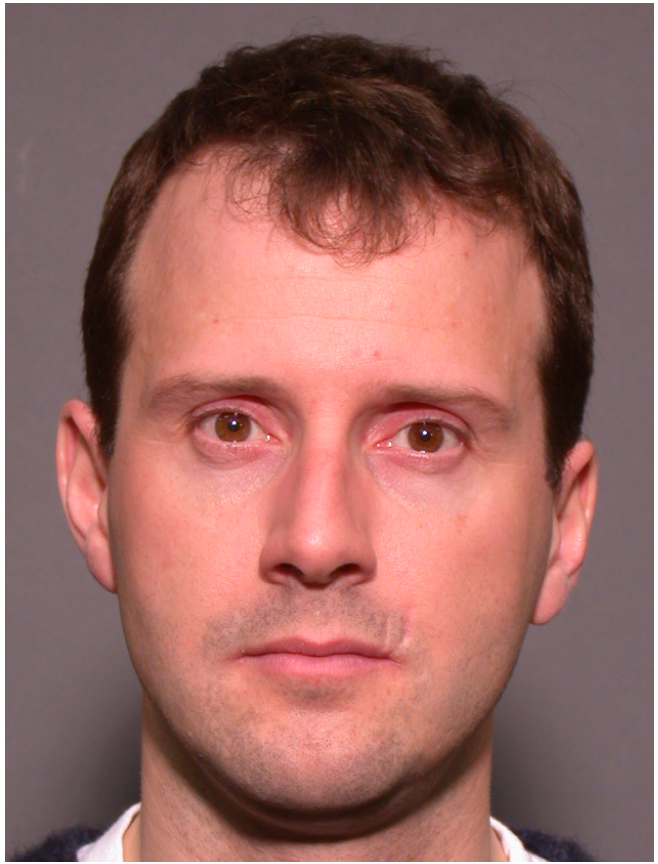
Low FPIR is not accessible

In a “closed” population (town, country):

- Low false positive rates cannot be achieved due to familial relationships
- Not expected with 10 fingerprints, and iris recognition



What's in a face?



How many biometrics here?

1 **Face**

2 **Irides + periocular**

3 **Skin texture** — https://patents.google.com/patent/US7369685B2/

4 **Head shape**

5 **Ears**

6 **Scars**

Human review: See ASTM E3149
Standard Guide for Facial Image Comparison Feature List for Morphological Analysis

7 **What other modalities could standardized augmentations of 2D face**

- Iris
- Short + long wave infrared
- Hyperspectral
- 3D

EVALUATING ACCURACY OF FACE QUALITY SCALAR ALGORITHMS

ONGOING BENCHMARKS

1. FRVT 1:1
Core Biometric Operation

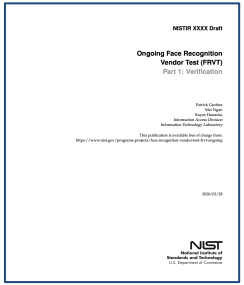
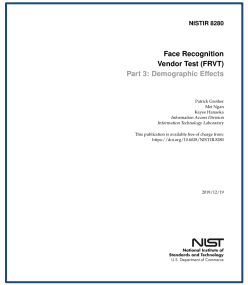
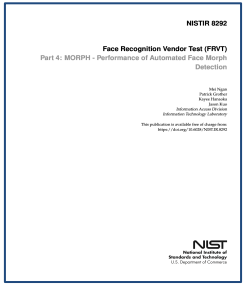
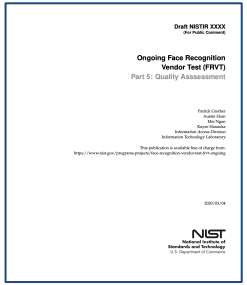
2. FRVT 1:N
Search Performance

3. FRVT Morph
Morphed Photo Detection

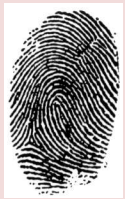
4. FRVT Quality
Automated Quality Assessment

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

Good, bad, wild, ugly, and lots beyond



distorted source e.g. scars on a fingertip

distortion in one or more steps of the process e.g. capture or compression

ISO Standard

Expression

Gaze

Too close

Pose Angle

- ISO's idea of "poor" images have better quality than (USA) border crossing practice
- ISO aspires to collect reference samples that are pristine, for storage in authoritative databases.

ISO* MUGSHOT+ WEBCAM LFW IJB-C LEGACY-CAM+ REFLECTEDx LUX = 0

Cooperative Enrollment where quality assessment is traditionally most useful at initial collection.

Non-Cooperative

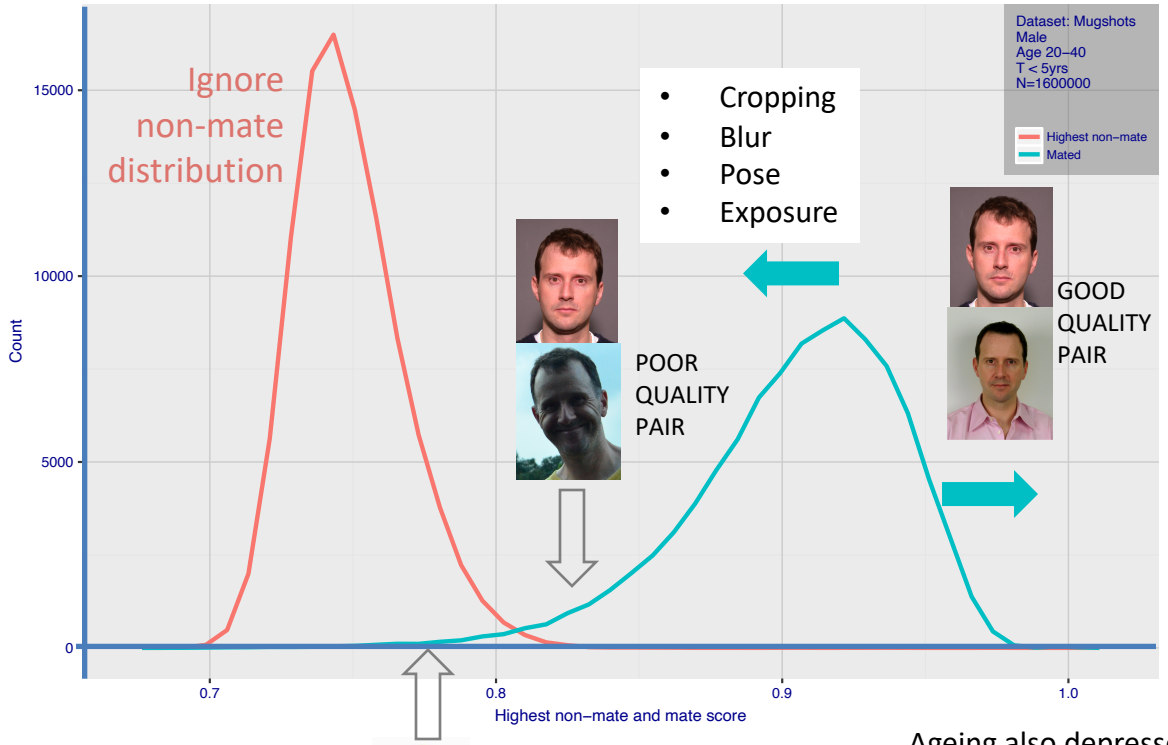
* http://webstore.ansi.org

\+ http://www.chicagonow.com/cta-tattler/2013/07/chicago-cops-use-face-recognition-software-to-nab-cta-mugger

x http://io9.com/hidden-faces-can-be-found-by-zooming-into-hi-res-photos-1491607189

Face Recognition Failures I

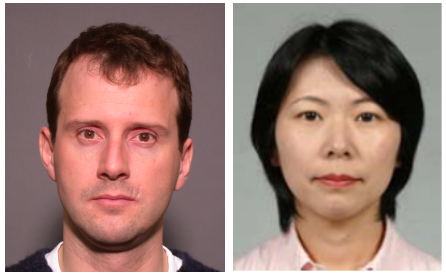
False Rejection



Quality as predictors of recognition success?

- A quality algorithm, F, operating on an image X_1 produces value
  - $Q = F(X_1)$ [1]
- A face recognition algorithm, C, samples to yield scores
  - $S = C(X_1, X_2)$ [2]
- Quality algorithms should predict genuine score, S, from X_1 alone
- By assuming that X_2 would be a canonical portrait i.e. a pristine image of the same subject
  - $Q \sim C(X_1, X_{PORTRAIT})$ [3]
  - i.e. quality assessment must be done "blind", targeting a hidden or virtual portrait image
    - cf. blind PSNR in image or video fidelity
  - Respects the ISO/ICAO specification as the gold standard for AFR.

FRVT Quality Evaluation Data



REFERENCE SAMPLES, X1

- Almost ISO compliant **immigration "application" photos**
- Num images: 535 329
- Num people: 535 329
- Percent female: 57.4%

VERIFICATION SAMPLES (Examples are very similar, not actual)

- **Inbound border crossing photos**
- Num images: 3 225 633
- Num people: 535 329
- Run quality algorithms on these images

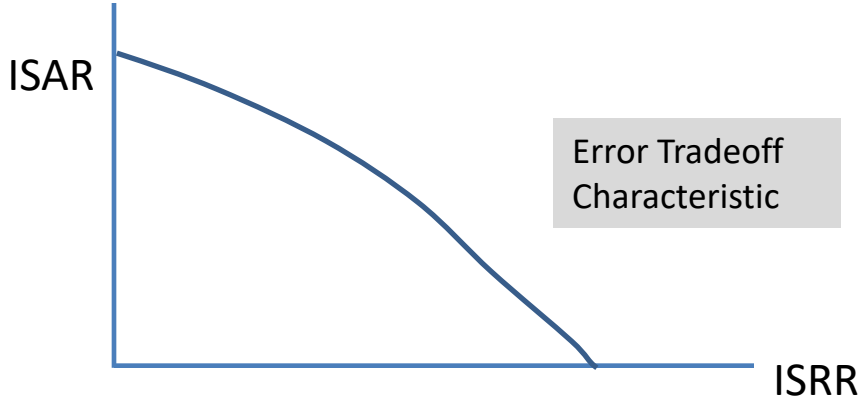
Dataset for quality evaluation:

- 3 225 633 comparison scores for each recognition algorithm
- 3 225 633 quality scores for each quality algorithm

Metric 2: QA as a predictor of recognition failure

X1	Q1	X2	Q2	MATCH SCORE	MATCH YES?
	?		94	0.97	TRUE
	?		97	0.91	TRUE
	?		74	0.89	TRUE
	?		57	0.85	TRUE
	?		29	0.81	TRUE
	?		68	0.72	FALSE
	?		32	0.65	FALSE
	?		27	0.57	FALSE

T



Incorrect Sample Rejection: Q < 55 but does match

Incorrect Sample Acceptance: Q ≥ 55 but does not match

FRVT Quality Assessment: Participation

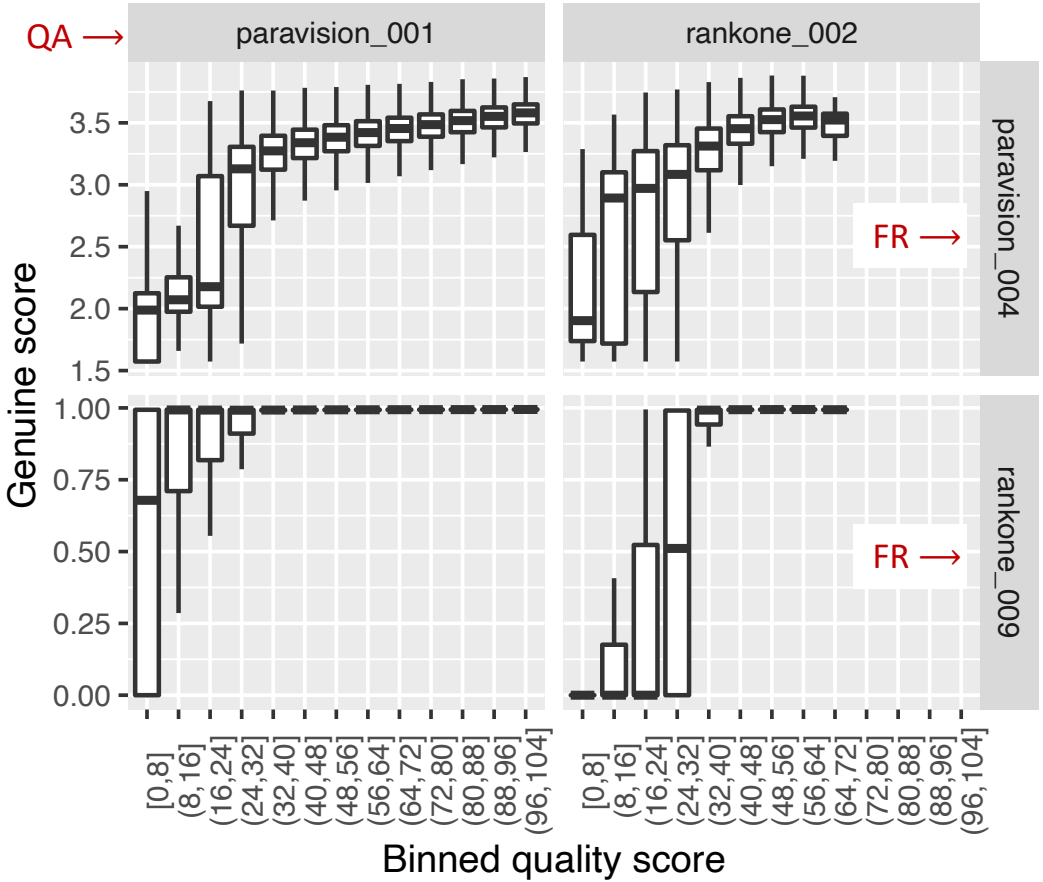
Quality algorithm developers choosing to participation

- China Electronics Import-Export (CN)
- Lomonosov Moscow State University (RU)
- Paravision (US)
- Guangzhou Pixel Solutions (CN)
- Rank One Computing (US) x3
- Universidad Autónoma de Madrid + Joint Research Center (EU)

Target recognition algorithms selected by NIST

- ceiec-003
- intsysmsu-001
- paravision-004
- pixelall-003
- rankone-008 and -009
- anyvision-004
- **imperial-002**
- innovatrics-004

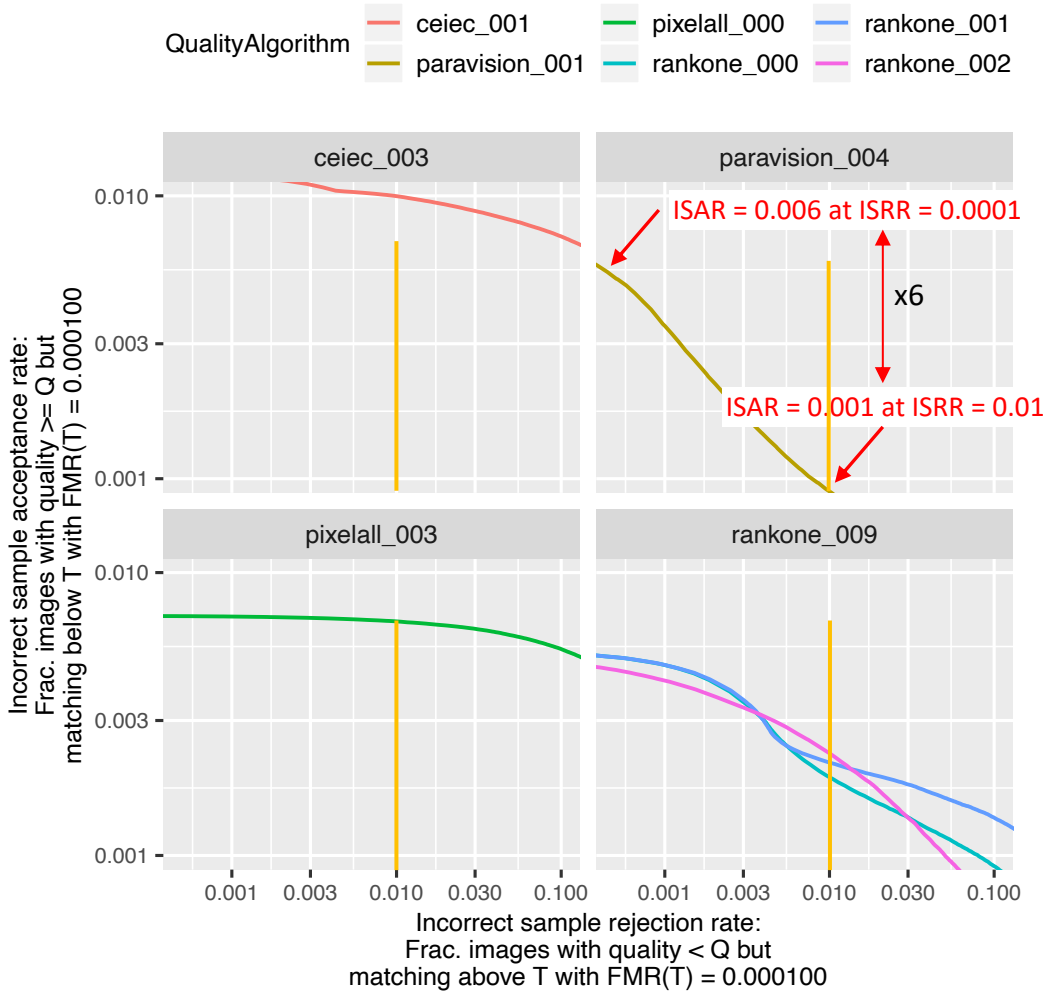
Result I: Suitability for use as “summary indicator”



- Monotonic medians
- Binned to 13 levels
- Variance sometimes high
- Within- vs. cross-developer

Result II: Suitability for use in image acceptance

- ISRR = 0.01 is an operationally tenable value (?)
- ISAR is it effective at improving accuracy?



FRVT Quality Evaluation

- Is open
- Is under development – new, and more, image sets
- https://pages.nist.gov/frvt/html/frvt_quality.html

THANKS

PATRICK.GROTHER@NIST.GOV

http://paddymondo.net/ISO_IEC_29794_5.pdf

http://paddymondo.net/ISO_IEC_24358.pdf