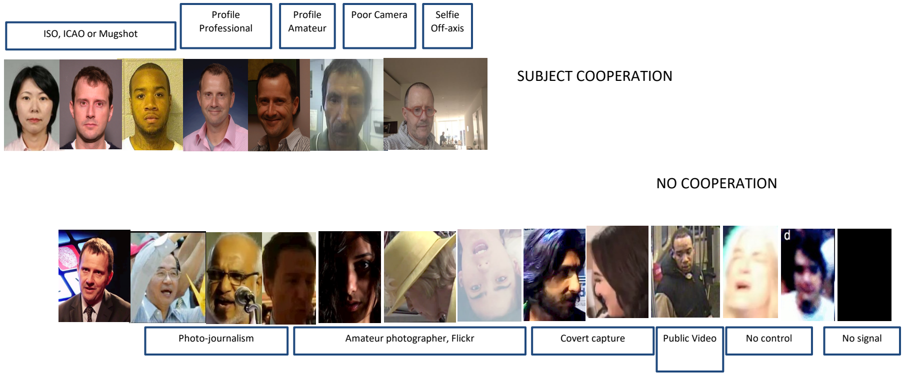
NIST Quality Evaluations ISO/IEC 29794-5

Patrick Grother

NIST
U. S. Department of Commerce

NIST/EAB/DHS OBIM Quality Workshop

November 18, 2021



14:45	Patrick Grother	**NIST FRVT FaceQuality – Vector Elements**
15:25	Patrick Grother	**Evolution of ISO/IEC IS 29794-5:202x**

ISO/IEC 29794-5 Face Image Quality

NIST

» Document is working draft

» Discussion in 2022-01 in SC 37 Working Group 3

» It is available to the public (but ISO's website prevents that)
  - https://www.iso.org/standard/81005.html
  - Instead today: http://paddymondo.net/ISO_Q.pdf

ONGOING BENCHMARKS

1. **FRVT 1:1** Verification
2. **FRVT 1:N** Search Performance
3. **FRVT Morph** Morphed Photo Detection
4. **FRVT Quality** Automated Quality Assessment

…

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: Use of Face Recognition in Paperless Travel
NISTIR XXXX Draft Ongoing Face Recognition Vendor Test (FRVT) Part 1: Verification	NISTIR 8271 DRAFT SUPPLEMENT Face Recognition Vendor Test (FRVT) Part 2: Identification	NISTIR 8280 Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects	NISTIR 8292 Face Recognition Vendor Test (FRVT) Part 4: MORPH - Performance of Automated Face Morph Detection	Draft NISTIR XXXX Ongoing Face Recognition Vendor Test (FRVT) Part 5: Quality Asssessment	NISTIR XXXX Ongoing Face Recognition Vendor Test (FRVT) Part 6A: Face recognition accuracy with face masks using pre-COVID-19 algorithms	NISTIR 8381 Face Recognition Vendor Test (FRVT) Part 7: Identification for Paperless Travel and Immigration
Last: 2021-10 Next: 2021-11-19	**Last: 2021-10** Next: 2021-11-19	Last: 2019-12-19 Next: 2021-12 est.	**Last: 2021-10** Next: 2021-10 est.	**Last: 2021-08** Next: 2021-11 est.	**Last: 2020-10** Next: 2021-11 est.	**Last: 2021-10** Next: 2021-12

…

https://pages.nist.gov/frvt/html/frvt1N.html

FRVT: New Benchmarks

ONGOING BENCHMARKS

1. FRVT 1:1
Verification

2. FRVT 1:N
Search Performance

3. FRVT Morph
Morphed Photo Detection

NOW

4A FRVT Quality
Automated Quality Quantification

FUTURE

4B FRVT Quality
Specific Image Defect Detection

FUTURE

5. FRVT Attack
Presentation Attack Detection

Q as predictor of false match

Nature of the problem

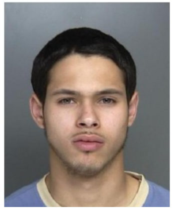
- Some false matches are due to quality problems
  - Example: overexposure
  - Example: common hair patterns on face
- Most false matches are due to biological similarity of the two faces.
  - Two people are involved!!
  - This occurs even in images that are high quality
- For this reason you cannot immediately evaluate a QA algorithm on its ability to predict FMR.
  - You'd have to isolate JUST those cases where false match is not obviously due to people being of similar appearance.

Evaluations of QA software

- Should be conducted to predict FNMR i.e. low mate scores.
- Mate scores are low due to: a) quality, b) ageing, c) injury or surgery or facial hair
  - Item b) can be excluded as a factor by test design
- This supports cooperative applications like passport issuance.
- In operations, quality is most useful when only one sample exists
  - Applying for an ID card at a new job.
  - Applying for a passport
- For that reason, quality evaluations should be conducted without pairwise quality combinations
  - Better to avoid: min(Q1, Q2), sqrt(Q1 Q2)

FRVT Quality Tracks

TRACK A
Q Summaries



SCALAR: Q = 98

DECISION: Y, Accept

BOX 0. QUALITY BENCHMARK

- One ”visa – border” dataset
- No longer use wild
- Extend to use new ”kiosk” dataset

TRACK B
Q diagnostics

BOX 1. QUALITY BENCHMARK

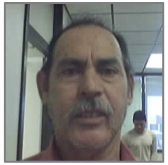
- Concept presented at the Nov Q Workshop
  - Publish 2021-11
  - Developer comment
- Algorithms to NIST 2022-01
- Align with ISO/IEC 29794-5

BOX 2. IMAGING VARIABLES THAT INFLUENCE ACCURACY

- Illumination adequacy + uniformity
- Exposure
- Focus, blur
- Resolution / Sp. Sampling Rate

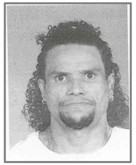
BOX 3. SUBJECT VARIABLES THAT INFLUENCE ACCURACY

- Head orientation (R, P, Y)
- Expression neutrality
- Sunglasses, face masks
- Motion blur
- No, or additional, faces



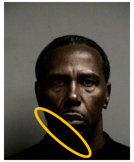
Two People

No People



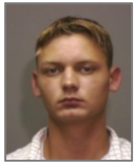
Noise

Over-exposure

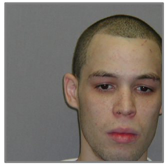


Under-exposure

Hot Spots



Mis-focus



Cropped

Non-frontal

ISO/IEC 29794-5 “capture-related” elements in clause 6.3

Quantity	Clause	Collection of reference samples for ID credentials	Collection of probe for instantaneous recognition	System enrolment, current or later creation of a reference, delayed recognition
Background uniformity	6.3.2		optional	optional
Illumination uniformity	6.3.3		optional	optional
First moment: Brightness	6.3.4		optional	optional
Second moment: Variance			optional	optional
Third moment: Skewness			optional	optional
Fourth moment: Kurtosis			optional	optional
Illumination non-overexposure	6.3.5		optional	optional
Illumination overexposure	6.3.6		optional	optional
Dynamic range	6.3.7		optional	optional
De-focus	6.3.8		optional	optional
Image sharpness	6.3.9		optional	optional
Motion blur	6.3.10		optional	optional
Edge Density	6.3.11		optional	optional
Compression	6.3.12	optional	optional	optional
Unnatural colour, colour balance	6.3.13		optional	optional
Camera lens focal length	6.3.14	optional	optional	optional
Camera-subject distance	6.3.15	optional	optional	optional

ISO/IEC 29794-5 “subject-related” elements in clause 6.4

Property	Clause	Collection of reference samples for ID credentials	Collection of probe for instantaneous recognition	System enrolment, current or later creation of a reference, delayed recognition
Eyes visible	6.4.2		optional	
Inter-eye distance	6.4.3			
Horizontal position of the face	6.4.5			
Vertical position of the face	6.4.6			
Pose – Yaw	6.4.7		optional	
Pose – Pitch			optional	
Pose – Roll			optional	
Expression neutrality	6.4.8		optional	optional
Mouth closed	6.4.9		optional	
Eyes open	6.4.10		optional	optional

Quality diagnostics vector

Black box algorithm, implementing NIST defined C++ API, supplied as compiled dynamically linked lib.so

Vector quality, **x**

- Real valued
- Named and defined elements (see right)
- Ultimately standardized as ISO/IEC 29794-5

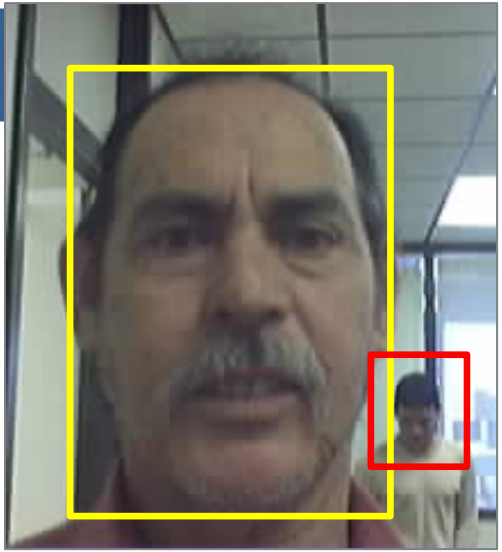
#	Name	Notes
1	Scalar Quality Value	As tested in FRVT now
2	Roll	Signed
3	Pitch	
4	Yaw	
5	Occlusion periocular	
6	Occlusion nose mouth	
7	Mouth open	
8	Eyes open	
9	Illumination adequacy	
10	Illumination uniformity	
11	Sharpness	Combines mis-focus, motion blur etc
12		
13		
14		

API

```
typedef map<string,double> quality;                // named real value

typedef vector<quality> Qvec;                      // quality vector

class Interface {
public:
   virtual ~Interface() {}

virtual FRVT::ReturnStatus
   initializeImageAnalysis(const std::string &configDir) = 0;

virtual FRVT::ReturnStatus
   scalarQuality(const FRVT::Image &, double &qvalue) = 0;

virtual FRVT::ReturnStatus
   vectorQuality(const FRVT::Image &, Qvec &diagnostics) = 0;

 static std::shared_ptr<Interface>
   getImplementation();
}
```

What else?

1. Eye coordinates?



Usage:

```
1.   diagnostics.size();                             // 2 faces
2.   diagnostics[0][“yaw”];                          // -2.2 degrees
3.   diagnostics[1][“pitch”];                        // 30.8 degrees
4.   diagnostics[0][“mouthopen”];                    // 0.1
5.   diagnostics[0][“illumination adequacy”];        // 4.7 bits
```

Face count

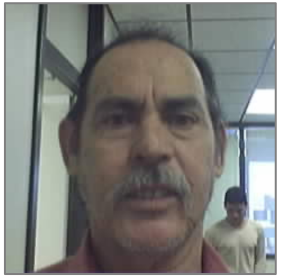
Task
- Count the number of faces in the image, including those of the subject, people in the background, on T-shirts, in photos on the walls behind, even if cropped.

Motivation
- In applications where one face is assumed, other faces can be detected instead of the intended one, leading to false negatives.

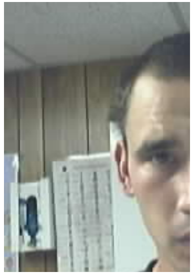
Software output
- A count of faces, and their locations.

Evaluation
- Runs on sets of images good images (N = 1)
- Run on sets of problematic images (N ≠ 1)

Metrics
- Confusion matrix
- Tabulate by class of image


2


0


1


1


2

Specific image defect: Non-frontal head orientation

NIST

Task
- Estimate the orientation of face (with respect to the camera):
- The head may not be close to the optical axis.

Motivation
- Head orientation different than the canonical frontal degrades accuracy

Software output
- Estimates of Roll, Pitch, and Yaw

Evaluation
- Run on images with known-by-design, or hand-coded, orientations

Metrics

Penalize estimators independently

- $F_{YAW}(\theta_{ESTIMATE} - \theta_{TRUTH})$
- $F_{PITCH}(\theta_{ESTIMATE} - \theta_{TRUTH})$ tolerant of definitional problem
- $F_{ROLL}(\theta_{ESTIMATE} - \theta_{TRUTH})$

With penalty perhaps $F(\phi) = 1 - \cos(a\phi)$ with scale factor "a"

- $a_{ROLL} > a_{YAW} > a_{PITCH}$
- values TBD

Yaw = +59 degrees
Pitch = 0 degrees
Roll = 0 degrees

Yaw = -37 degrees
Pitch = +4 degrees
Roll = +1 degrees

Yaw = -90 degrees
Pitch = 0 degrees
Roll = 0 degrees

Yaw = -22 degrees
Pitch = +3 degrees
Roll = -18 degrees

Illumination uniformity

Task
- Quantify whether face is lit uniformly

Motivation
- Sufficient illumination non-uniformity will produce false negatives

Software output
- Measure of non-uniformity

Evaluation
- Runs on sets images with varying directional illumination

Metrics
- Report pairwise statistics of ground-truth and measured value

Eye glasses present

Task
- Detect eye glasses and sunglasses

Motivation
- False positives from glasses
- False negatives and false positives from glasses

Software output
- Boolean presence indicator
- ?? Measure frame thickness as in 39794-5:2019 Clause

Evaluation
- Runs on sets images with and without glasses

Metrics
- Confusion matrix
- Summary measure: β FNR + (1-β) FPR with high β

Occlusion Periocular Region

Task
- Quantify how occluded the periocular region is (by hair, glasses, sunglasses)

Motivation
- Occlusion can impede detection and elevate FNMR

Software output
- Fraction of region that is occluded

Evaluation
- Runs on sets of images with various levels of occlusion

Metrics
- Report pairwise statistics of ground-truth and measured value

Q1: How to define periocular region

Q2: How to handle transparent glasses

Occlusion Nose and mouth

Task
- Quantify how occluded the nose and mouth region are

Motivation
- Occlusion can impede detection and elevate FNMR

Software output
- Fraction of region that is occluded

Evaluation
- Runs on sets of images with various levels of occlusion

Metrics
- Report pairwise statistics of ground-truth and measured value

Q1: How to define nose-mouth-chin region

Q2: Do we care about nose occlusion at all?

Mouth open

Task
- Determine if the mouth is open, as prohibited in standards

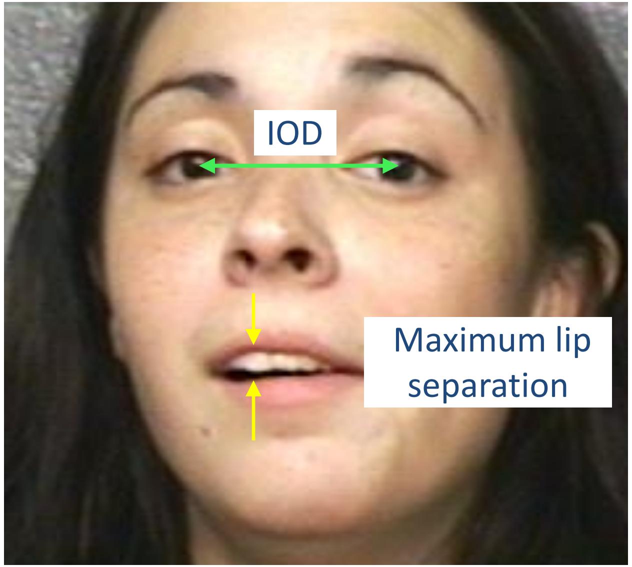
Motivation
- False positives from glasses
- False negatives and false positives from glasses

Software output
- EITHER κ = maximum distance between lips / estimated interocular distance
- OR a mouth-openness measure

Evaluation: Using images for which ground truth is manually known:
- Run on images with mouth closed
- Run on images with mouth open

Metrics
- EITHER report joint distribution of $\kappa_{ESTIMATE}$ and κ_{KNOWN}
- OR tradeoff as a function of openness threshold, compute
  1. Rate of false assertion of mouth being open
  2. Rate of missed detection of mouth being open

YES NO



Specific image defect: Under-exposure

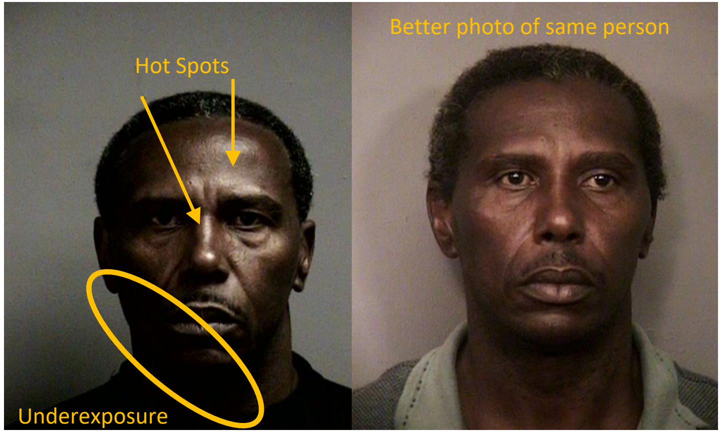
Task
- Detect underexposure of the face in an image

Motivation
- Under exposure drives higher false negative rates
- Likely induces demographic dependence therein

Software output
- An entropy measure in the face region

Evaluation
- Run on perfect images and those with a wide range of under-exposure

Metrics
- Report joint distribution of $H_{ESTIMATE}$ and H_{TRUTH}
- Summarize with RMS

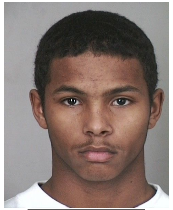


Source: NIST Special Database 32 aka "MEDS", subject S171

H = 3

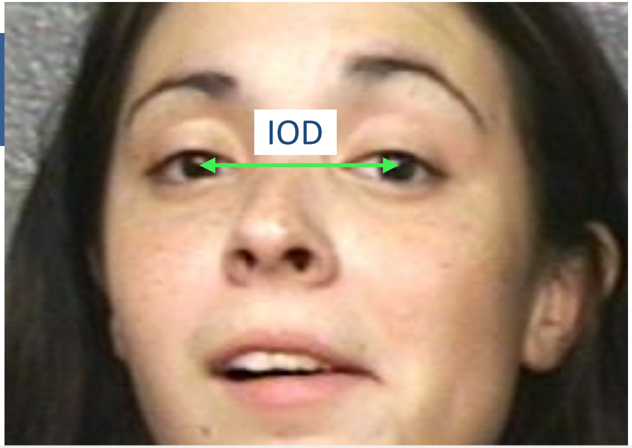
H = 2.4

H = 4.1



H = 7.1

Interocular Distance



Task

- Compute IOD as measure of spatial sampling rate
- Convert IOD to a higher-is-better quality measure on [0,1]

Motivation

- Small faces elevate FNMR (and FMR).

Software output

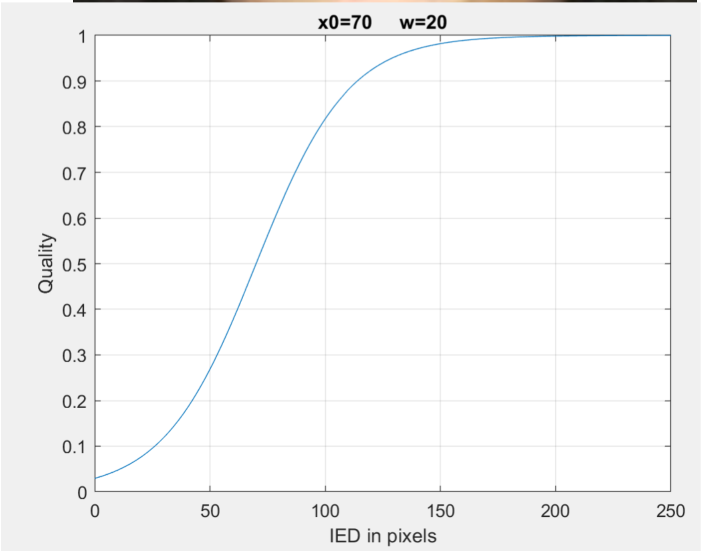
- Sigmoidal “conformance” value

Evaluation

- Runs on sets of images with various IOD

Metrics

- Report pairwise statistics of ground-truth and measured value



Misplacement and size

Task
- Determine whether the subject is positioned properly in the field of view

Motivation
- Mispositioning can impede detection

Software output
- Location
- Location non-conformance $SIGMOID(X\text{c}\ , 0.45A, 20, 100)$ where A is the image width, Xc is midpoint between eeyes

Evaluation
- Runs on sets images with cropping and margins variations

Metrics
- Report pairwise statistics on estimated vs. ground truth




Q Assessment Formalism

Scalar Quality

- Compute quality from an image
  - $q_1 = F(x_1)$
- Face recognition mate score
  - $s = C(x_1, x_2)$
- Evaluate quality as predictor of comparison
  - Low q $\longrightarrow$ Low s   using error-vs-reject
  - Reject only low percentages (say < 5%)

Vector Quality

- Compute image analysis vector
  - $v_1 = G(x_1)$
- Evaluate each element separately against ground truth
  - Is the yaw estimate accurate?
  - Is the crop measure close to correct?
  - Are eye glasses detected correctly?
  - …
- Can we evaluate that vector by relating it to mate scores?
  - Next slide

Evaluating the image quality vector

Compute image quality vector

- $v_1 = G(x_1)$

Function to be provided by developer to NIST evaluation AND commercially as it has obvious operational relevance: "what's wrong with an image"

If the ISO/IEC 29794-5 is mandating measurement of important image properties, the vector should relate to mate scores. So possibilities:

- $R(v_1) \sim S = C(x_1, x_2)$

Un-necessarily restrictive. Bad Idea!

- $R(v_1, v_2) \sim S = C(x_1, x_2)$

Function to be provided by developer for NIST evaluation.

It maps vectors from *both* images to model the score. Why?

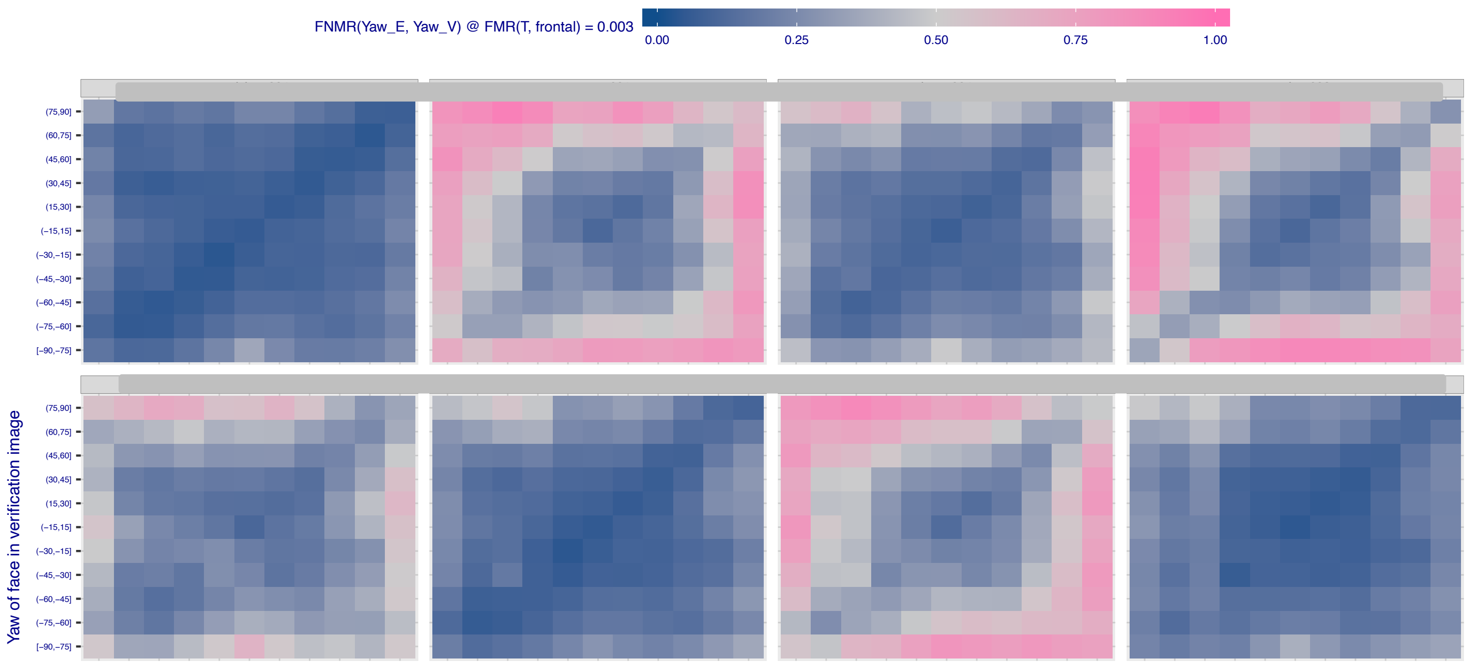
- To demonstrate that the vector elements do relate to false negative possibility
- To reveal what elements are salient c.f. random forest

Because the function operates on both images it has no operational relevance.

```
virtual FRVT::ReturnStatus
        QualityScorePredictor(const Qvec &v1,
                              const Qvec &v2
                              double &predictedscore) = 0;
```

Motivation for $R(v_1,v_2)$

Element Name	v_1	v_2	Mate score dependence	
…				
Yaw	-30	-35	score ~ $\lvert \theta_1 - \theta_2 \rvert$	
IOD	40 px	90 px	score ~ $\min(IOD_1, IOD_2)$	
Occlusion eyes	50%	0%	score ~ $\max(O_1, O_2)$	
…				



Pose Invariance

FNMR vs. Yaw-Yaw at FMR = 0.003

On [-90,90]

Following in the NFIQ footsteps

NFIQ 2.1

- Developed US-EU team, 2010 - 2017 ...
- Input: Greyscale 500ppi plain impression
- Features: Handcrafted
- Classifier: Random forest classifier
- Output: 0 ≤ Q ≤ 100
- Trained to predict some function of genuine and impostor scores from commercial algorithms applied to operational optical fingerprint images
- Standardized as ISO/IEC 29794-4:2017
- Evolution discussed 2021-06
  - https://eab.org/events/program/248
- Contact: nfiq2@nist.gov

Face Equivalent

- Input: JPEG ...
- Features: ?
- Classifier: ?
- Output: 0 ≤ Q ≤ 100
- Trained to predict some function of genuine and impostor scores from commercial algorithms applied to operational face photos
- To be standardized as ISO/IEC 29794-5

Mapping v to q

Compute image quality vector

- $v_1 = G(x_1)$

```
virtual FRVT::ReturnStatus
   vectorQuality(const FRVT::Image &, Qvec &diagnostics) = 0;
```

Evaluate goodness of vector via

- Does $R(v_1, v_2)$ predict scores from some large set $S = C(x_1, x_2)$

```
virtual FRVT::ReturnStatus
            QualityScorePredictor(const Qvec &v1, const Qvec &v2, double &predictedscore) = 0;
```

Train function R via

```
virtual FRVT::ReturnStatus
            QualityScoreTrain(const vector<Qvec> &v,
                              const vector<double> &actualscores,
                              // some output class here
                              ) = 0;
```

FRVT: Coming changes

ONGOING BENCHMARKS

- **1. FRVT 1:1** Verification
- **2. FRVT 1:N** Search Performance
- **3. FRVT Morph** Morphed Photo Detection
- **4A FRVT Quality** Automated Quality Quantification

FUTURE

- **4B FRVT Quality** Specific Image Defect Detection
- **5. FRVT Attack** Presentation Attack Detection

New benchmarks

1. Extended benchmark on quality checks
2. Ongoing benchmark on presentation attack

New reports

1. Updates for new algorithms
2. Report on ability to disambiguate twins
3. Add new challenging datasets
4. Traceability
5. Demographics
    a) Summary equity measures
    b) Document developer improvements
    c) Update report, split into 1:1 and 1:N
    d) Add new datasets to 1:N

CURRENT PRODUCTS

UPCOMING (Part 3b) · UPCOMING (Part 8)

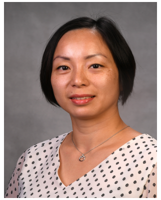
Part 1: Performance of 1:1 Verification Algorithms	Part 2: Performance of 1:N Identification Algorithms	Part 3: Demographic Effects in Face Recognition	Part 3b: Summarizing Demographic Differentials	Part 4: Performance of Morph Detection Algorithms	Part 5: Performance of Image Quality Assess. Algorithms	Part 6: Performance of Face Recognition with Face Masks	Part 7: Use of Face Recognition in Paperless Travel	Part 8: Performance of Face Recognition on Twins
								
Last: 2021-09-08 Next: 2021-10-08	Last: 2021-09-21 Next: 2021-10-12	Last: 2019-12-19 Next: 2022-01 est.	Last: 2021-Q4	Last: 2021-07-27 Next: 2021-10 est.	Last: 2021-08-11 Next: 2021-11 est.	Last: 2020-09-08 Next: 2021-10 est.	Last: 2021-07-13 Next: 2021-10-11	Last: Next: 2021-10

THANKS!

PATRICK.GROTHER@NIST.GOV

FRVT@NIST.GOV

Patrick Grother



Kayee Hanaoka

Austin Hom

Mei Ngan

Joyce Yang