Performance of Biometric Sample Quality Measures

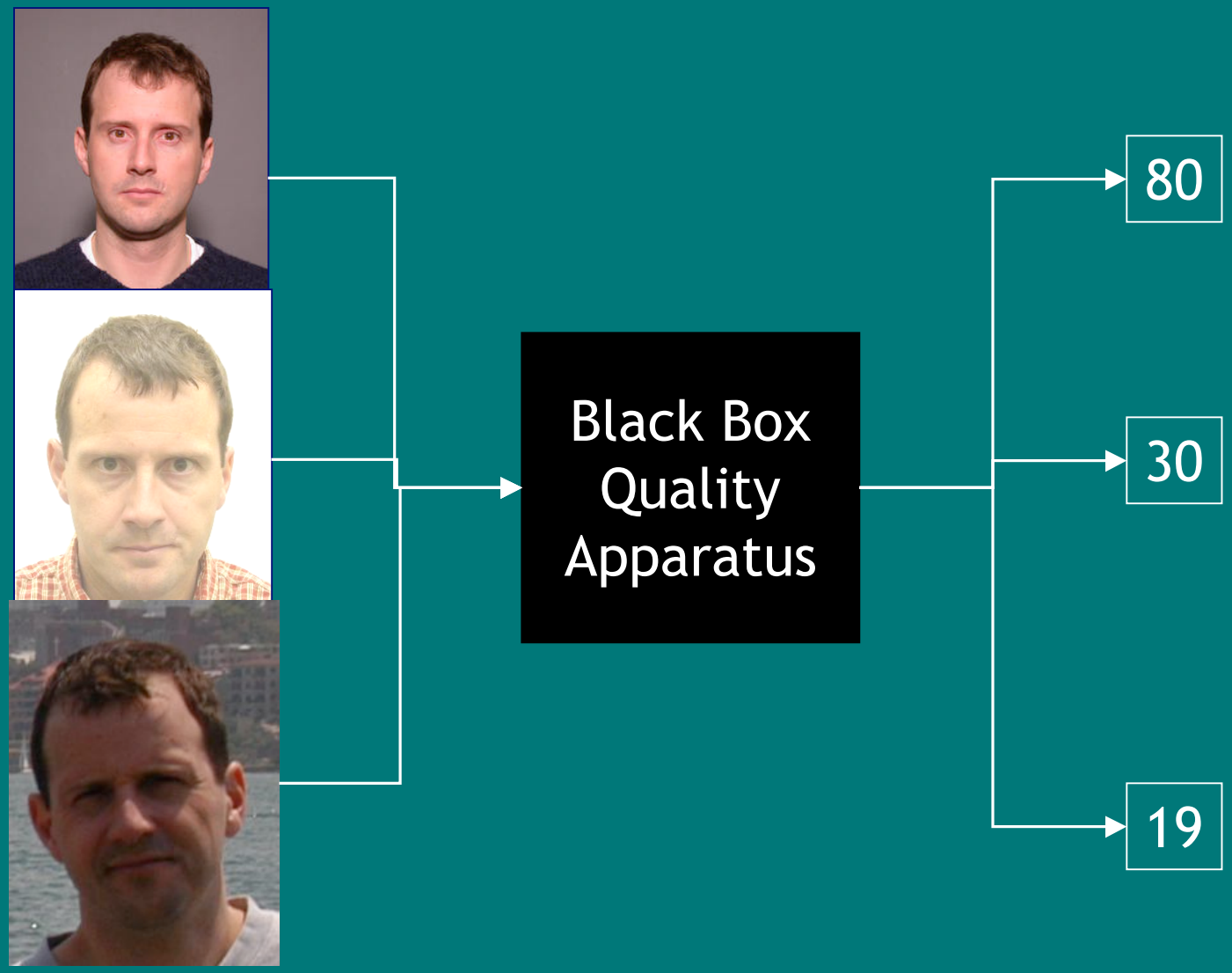
Definition, Utility, and Performance

Patrick Grother + Elham Tabassi
National Institute of Standards and Technology

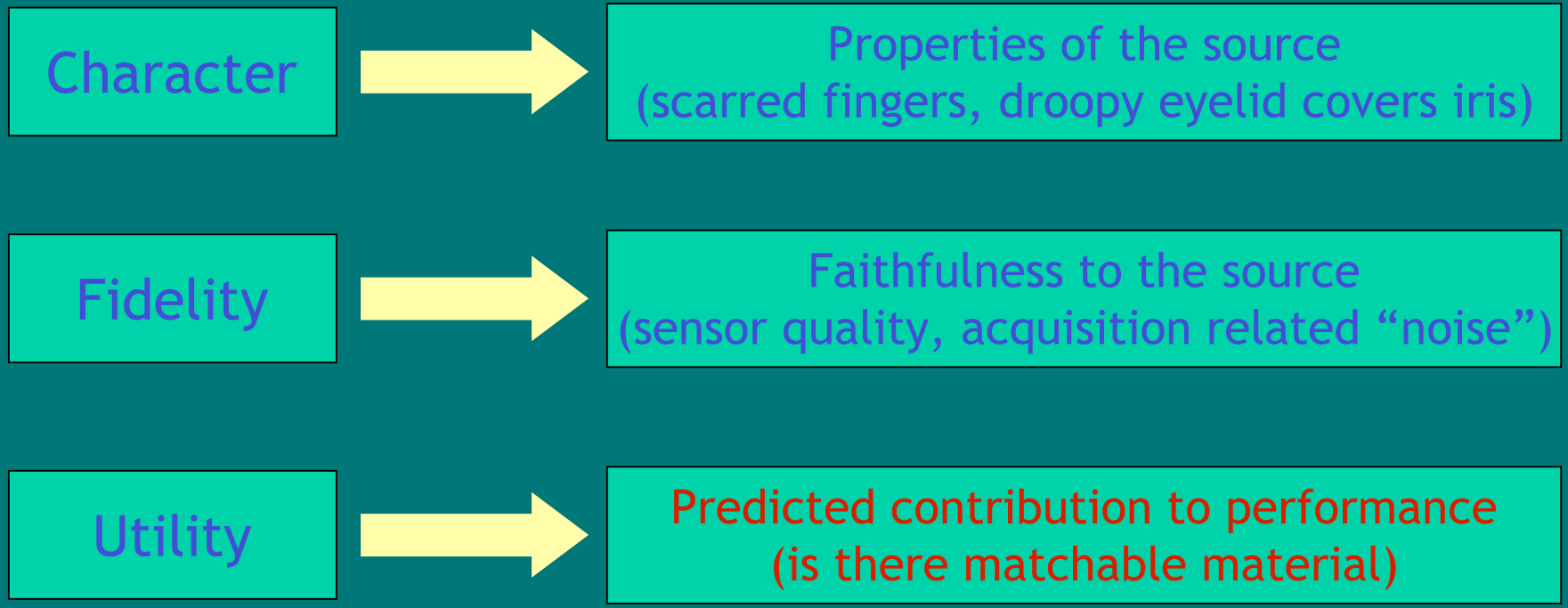
patrick.grother@nist.gov

AAAS, St Louis, February 20, 2006

Quality is Just a Scalar!



Three Kinds of Quality



Fingerprint Quality

Q = 2

Q = 4

	1	2	3	4	5
	BAD				GOOD

NIST Fingerprint Image Quality

feature extraction: computes image content and fidelity characteristics and results in an eleven dimensional feature vector.

neural network: classifies feature vectors into five classes of quality based on various quantiles of the normalized match score distribution.

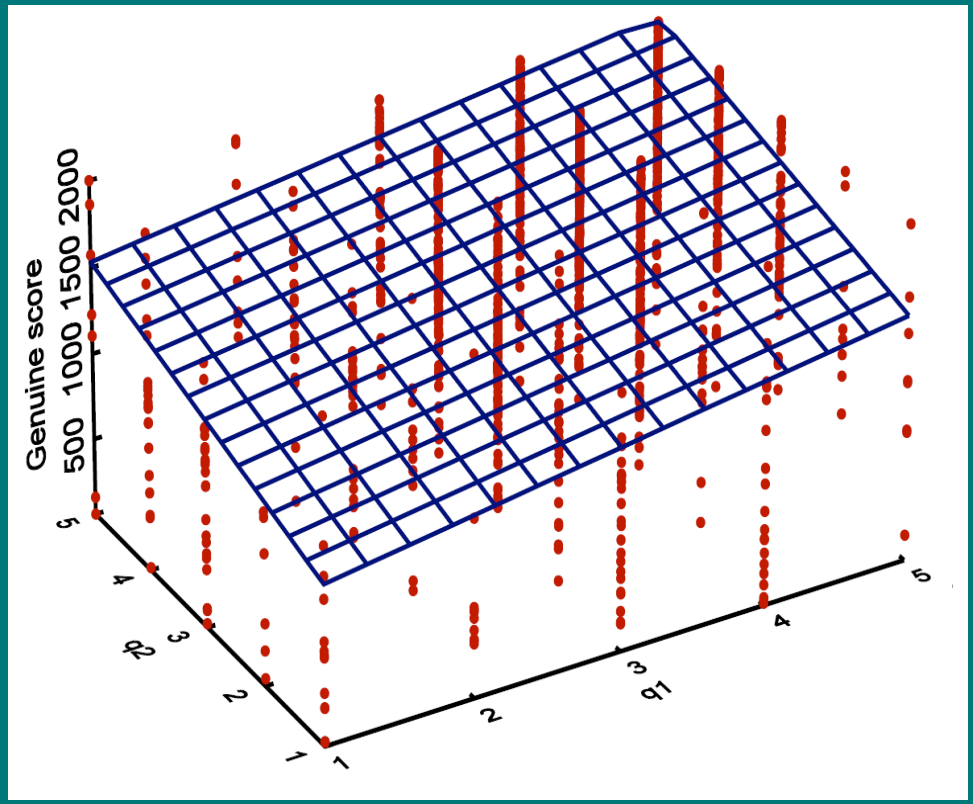
Uses of Quality Numbers

- Conditional reacquisition
  - Acceptance for enrollment
    - For credential issuance (visa, passport, access card, PIV)
  - Acceptance for verification
    - Of the samples just captured which one to send for matching?
    - Or acquire still more?
  - Acceptance for identification
    - Is the subject offering a poor sample deliberately?
- Initiate invocation of special processing or matching algorithms
- Quality directed fusion
- Quality Monitoring
  - Are some biometric field locations giving low quality?
  - Only in the evening?

Quality Values vs. Defect Bits

- Quality is a summary of bad (or good) traits
- But defect detection is more specific
  - Is the fingerprint image smudged?
  - Is the face image non-frontal?
  - Is the eyelid three-quarter closed?
- Knowledge of defect allows
  - Instruction to user

Matching Involves Two Samples



When Q1 and Q2 are both high the matching score is high.

Furthermore function is monotonic … BUT

when the enrollment sample is of good quality and better than that of the verification sample, the search sample's quality is sufficient to predict performance.

Why Quality as a Performance Predictor is Difficult

Authentication Score, S = F(Sample1, Sample2)

Quality value, Q1 = Q(Sample1)

Quality value, Q2 = Q(Sample2)

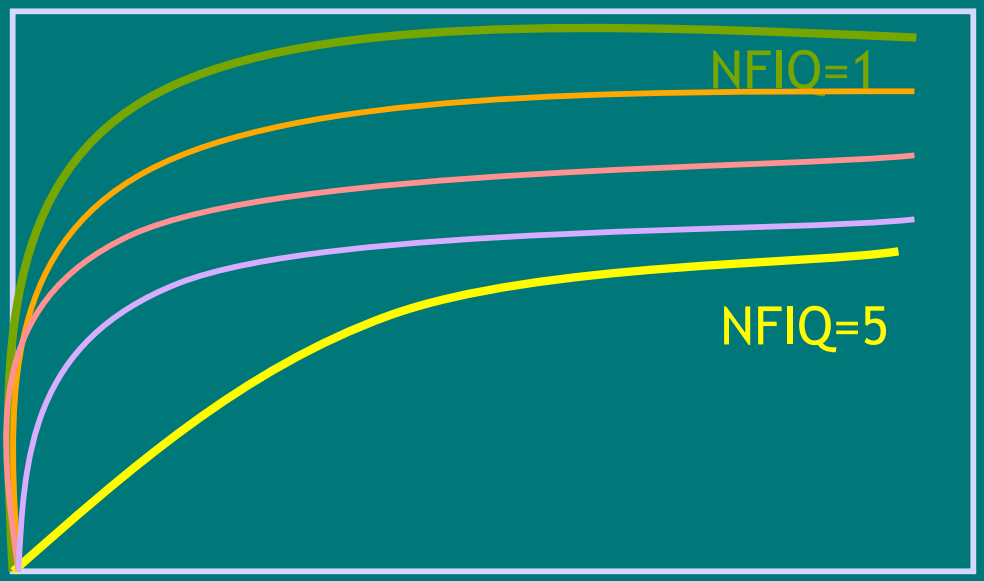
Score Estimate, E = P(Q1, Q2)

Quality and Performance

poor quality samples result in low performance

BLACK BOX QUALITY APPARATUS

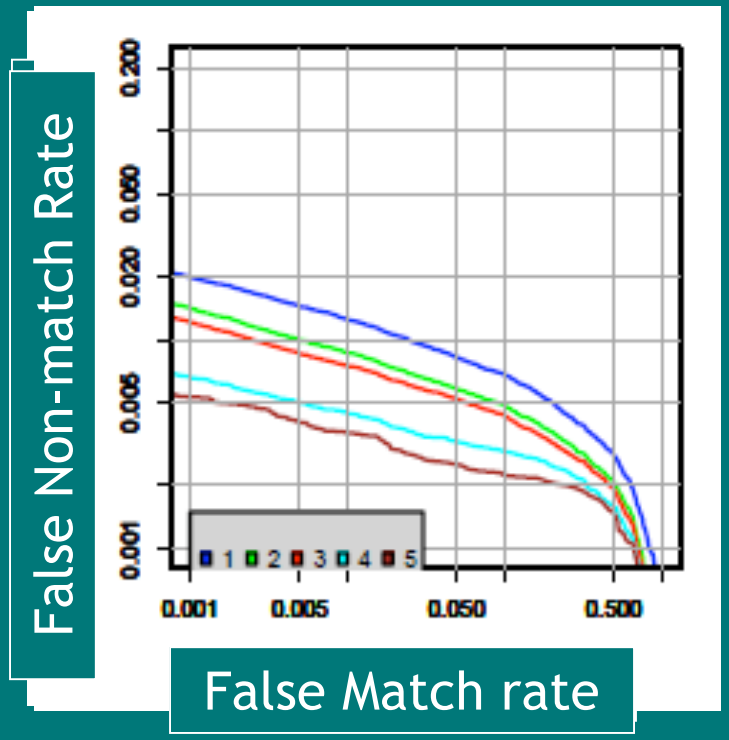
quality number =5



DET Curves

- Testing Case A
  - Enrol samples of quality X
  - Attempt authentication with samples of quality X
- Testing Case B
  - Enrol samples with quality ≥ X
  - Attempt authentication with samples of any quality

Case A vs B

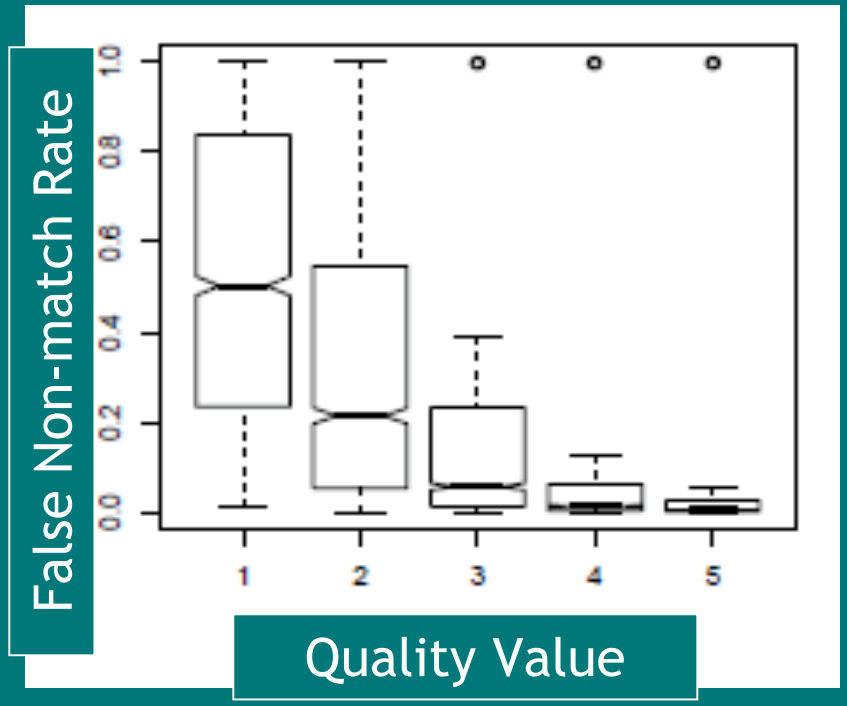


Q1 = Q2 = X

~~Q1 >= X~~

Q2 unconstrained

Quality vs. False Non-match Rate

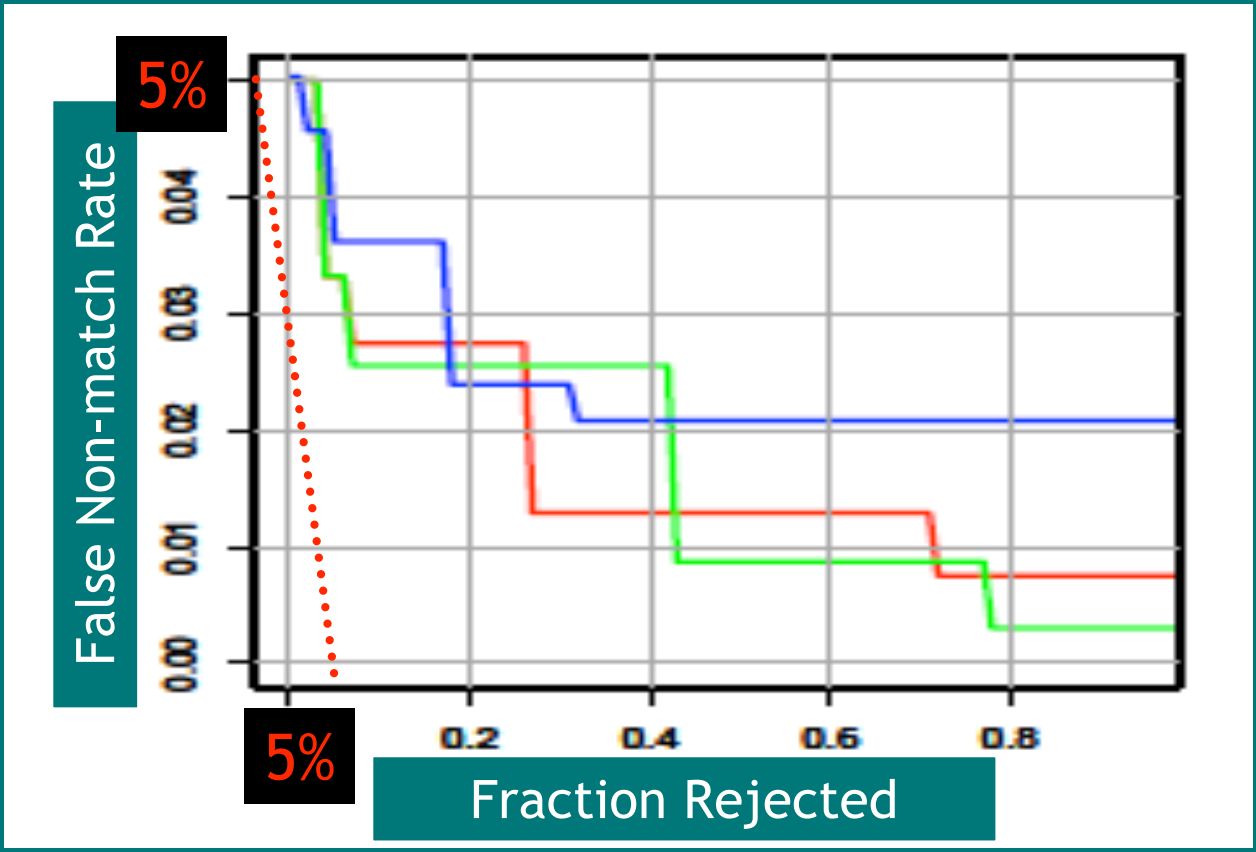


- Increasing quality gives lower false rejection rates
- Statistically distinct levels of performance
  - Five levels, not twenty

Error vs. Reject Metric

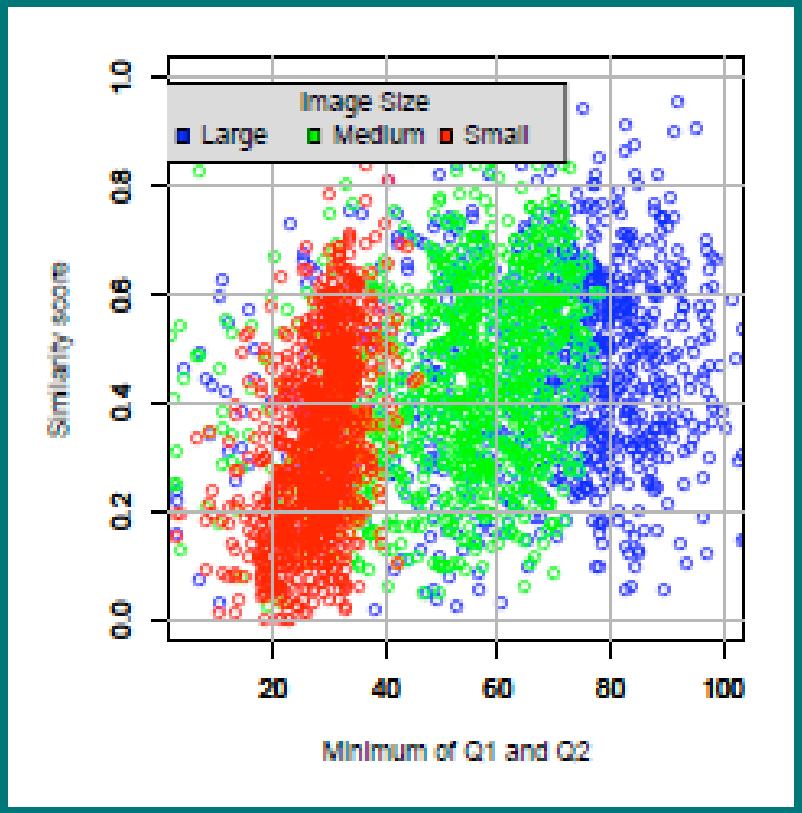
- Match two samples, get score S
- Compute quality of enrollment sample, Q1
- Compute quality of authentication sample, Q2
- Compute Q = min(Q1, Q2)
- Compute FNMR at some reasonable threshold t
- Sort (Q, S) on Q in decreasing order of quality
- Discard fraction x of lowest quality pairs
- Recompute FNMR
- Plot FNMR(x)

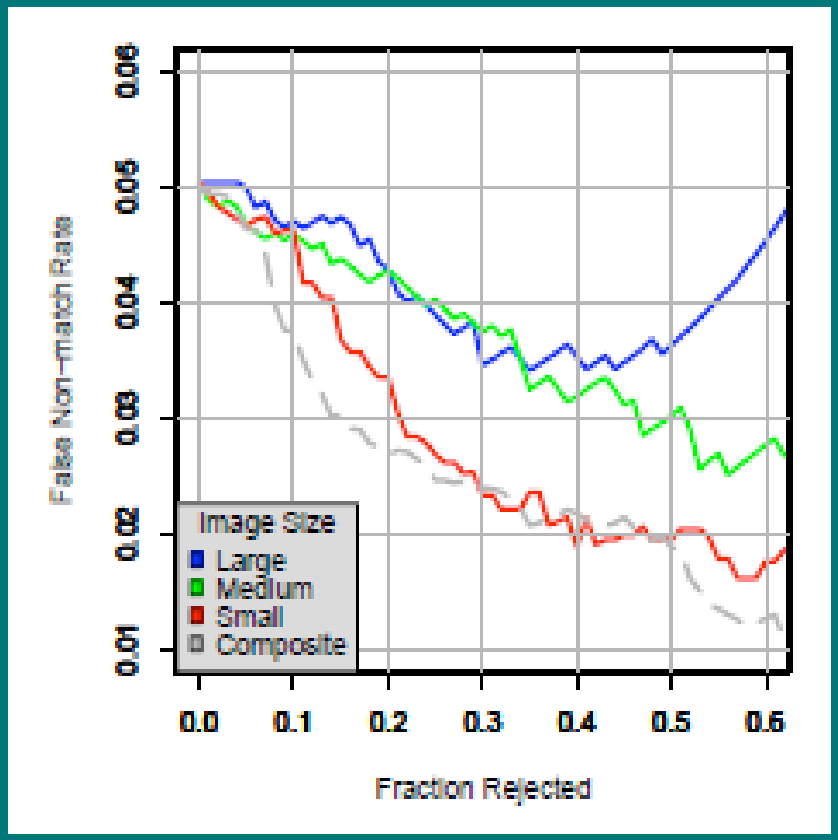
Evaluation of a Quality Measure



Existing quality measures are NOT perfect predictors of the worst case matching scores.

What Evaluation Data?





Conclusions

- Biometric quality assessment is an operationally important
- Quality assessment is difficult
- Relatively under-researched
- Quality measures can be evaluated if they're supposed to predict performance
  - In large scale matching trials
- Primary target should be false non-match rate
- Quality can usefully be represented as an integer
  - More statistically separable levels are better
- Quality measures likely to perform better if they reflect sensitivities of the matcher
- Matcher dependence is OK

Thank you

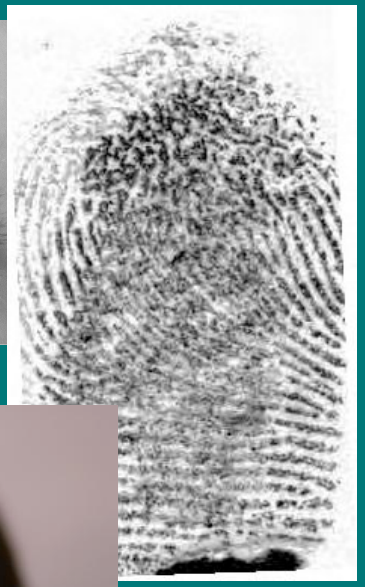
patrick.grother@nist.gov

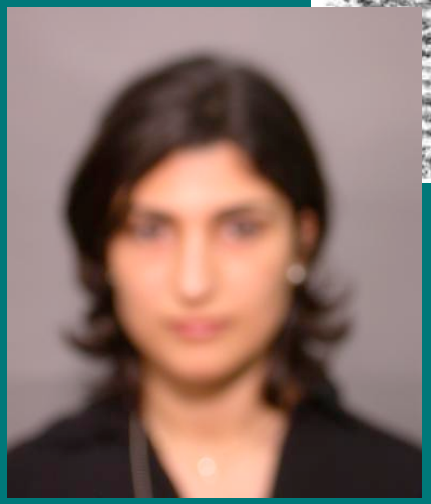
Quality Development cf. Pharmacology!

Patient	⟷	Biometric System
Drug (given to patient)	⟷	Sample (given to system)
Combination therapy	⟷	≥ 2 samples involved in matching
Efficacy (will it work)	⟷	Performance (will it match)
Clinical trial	⟷	Biometric test
Toxicity (poisons the patient)	⟷	Low quality samples produce false acceptances and duplicate enrollments (poisons the matcher)
Drug discovery (molecular targets)	⟷	Discover failure modes and sensitivities of the matcher
Goal is to tailor drugs and to narrow clinical trials	⟷	Goal is to only collect samples suited to matcher
Discovery methods follow decades of therapeutic use	⟷	Image quality and failure analysis follows years using images in biometric systems

patrick.grother@nist.gov

NIST Biometric Quality Workshop





March 8-9
NIST
Gaithersburg
Maryland

http://www.itl.nist.gov/iad/894.03/quality/workshop/

patrick.grother@nist.gov

NIST Biometric Quality Workshop

Applications
- What are the use-cases, business-cases and economics?
- Are there applications beyond sample re-acquisition, quality assurance and survey, detection of evasion, and fusion?

Capabilities
- Can quality measurements adequately select the best sample from a stream?
- Can real-time measurements of quality be used to reduce FTE and FTA?
- What can be achieved by sensor design alone?

Standardization
- Does conformance to existing data format standards guarantee quality?
- Would standard quality corpora be useful? Or standard reference algorithms?

Modalities
- What methods exist for assessment of face, finger, iris and speech quality?
- To what extent does good design guarantee quality?
- Does multimodal, multi-sample, or multi-sensor acquisition solve quality problems?

Evaluation
- Should quality be predictive of recognition performance, and evaluated as such?

Research and Development
- What research is being done?
- Is the amount of research commensurate with its operational importance?
- Should research funding go into quality-by-design? Quality measurement? Or both?

patrick.grother@nist.gov

Quality is not about Human Perception

Quality depends how close you look

The Core Biometric Function

matched with

= 0.87

Genuine Comparison

matched with

= 0.03

Impostor comparison