Biometric Quality

The push towards zero error biometrics

NIST Forensic Symposium

Elham Tabassi
Image Group
Information Access Division
Information Technology Laboratory
Friday, November 30, 12
Open set 1:N Search
1:N Latent Search

Latent image

Search Template

Features

Candidate List

Bob  1278
Alice  357
Mike  267
Lisa  245
Matt  209

FNIR (Miss Rate)

FPIR (False Alarm Rate)

FNIR (Miss Rate)

FPIR (False Alarm Rate)

N template Enrollment Database

Template
Quality of latent, Quality of the exemplar and the size of the overlap area between the two play significant role in accuracy and reliability of the conclusion. If the exemplar is not of sufficient quality, the conclusion should be Inconclusive.
Need for numerical evaluation of quality

To supplement the fingerprint examination process by one that has a statistical model, supported by appropriate databases for calculating numerical measures of weight of evidence
  - Perhaps based on automatic latent recognition algorithms
    - Some of VEO or NV were successfully identified by latent comparison algorithm. (NIST ELFT- EFTS).

Resolve the variability among the examiner’s value determination
  - For 356 latents, unanimous value determination achieved 43%.
  - 85% of NV decisions and 93% of VID decisions were repeated by the same examiner after a time gap while only 55% of VEO decisions were repeated. (Ulery et. al – PNAS 2011).

While acknowledging the overall reliability of the conclusions of majority of fingerprint comparisons performed over the past century, and their contribution to the criminal justice system
Lights-out Latent search

Quality assessment for error suppression

» Quantify information content of latent
  – Suitable for automatic feature markup or manual?

» Quantify quality of the reference print

» Reliability of latent to reference comparison
  – Sufficient overlap area?
Quality assessment for error suppression

Reference print

» Quantitative assessment of utility of the friction ridge print.
  – Is this suitable for matching?
  – NFIQ, others.

» By numbers:
  – 34.5% of operational IAFIS exemplars are of low quality (NFIQ 3,4,5).

Crime scene Latent print

» Quantitative assessment of information content of the latent.
» Latents recovered from crime scenes are often limited in size, of poor quality, distorted and affected by interference from the substrate.
» A growing body of literature questions scientific foundation and transparency of the evaluation of the weight of evidence associated with any particular fingerprint comparison
» Recent related work
  – On latent fingerprint Quality Yoon, Liu, Jain, 2012
  – Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm, C. Neumann, 2011
What is `quality’?

`standard’ definition

**quality**
the degree to which a biometric sample fulfils specified requirements for a targeted application
NOTE: Specified quality requirements may address aspects of quality such as focus, resolution, etc. Implicit quality requirements address the likelihood of achieving a correct matching result.

**quality score**
a quantitative expression of quality

**utility**
the observed performance of a biometric sample or set of samples in one or more biometric systems
NOTE: The character of the sample source and the fidelity of the processed samples contribute to—or similarly detract from—the utility of the sample
NOTE: Utility may combine performance measures such as FMR, FNMR, failure to enrol rate, and failure to acquire rate

Predictive of performance

A biometric quality assessment method derives a numerical quality value from an input biometric sample. The quality value is related to the biometric error rates that are likely to be realized when the sample is matched.
NIST fingerprint image quality (NFIQ 1.0)

» NIST developed NFIQ in 2004
  – Open source, publicly available
» Has become the de-facto standard
» Key innovation: quality as a rank statistic for performance
» NFIQ is a machine learning algorithm
  – Exploratory variables: image properties (minutiae, ridge clarity)
  – Response variable: separation of genuine and impostor comparison
» **feature extraction**: computes appropriate signal or image fidelity characteristics and results in an 11-dimensional feature vector.

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

» **quality number**: an integer value between 1 (highest) and 5 (poorest).
NIST Fingerprint Image Quality

NFIQ’s 5 levels of quality are intended to be predictive of the relative performance of a minutia-based fingerprint matching system. NFIQ=1 indicates high quality samples, so lower FRR and/or FAR is expected.
NFIQ=5 indicates poor quality samples, so higher FRR and/or FAR is expected.
NFIQ – feature vector

1 total # of minutiae
2 # of min. with q ≥ 0.5
3 # of min. with q ≥ 0.6
4 # of min. with q ≥ 0.7
5 # of min. with q ≥ 0.8
6 # of min. with q ≥ 0.9

7 size of foreground
8 quality zone 1
9 quality zone 2
10 quality zone 3
11 quality zone 4

NIST Minutiae detector (mindtct of NBIS distribution) has been used for feature extraction.
NFIQ 1.0 – training

training: 3900 images of flat index fingers and thumbs

A full similarity matrix of the training set is needed to compute the output class of neural network.

quality number \{1,2,3,4,5\}
1 is the best and 5 is the poorest
Separation of genuine and impostor distribution
NFIQ – rank statistic for performance
NFIQ 1.0 – test of time

+

» Novel definition of biometric quality
  – performance related
  – accepted by the community

» Interoperability
  – uniform interpretation
  – tuned to a class of matcher

» Open source

» Extensively examined
  – by NIST and others
  – tools for quality summarization, slap, ...

–

» Aging
  – recognition technology has advanced since 2004.

» Efficiency
  – ~300 msec per image - not fast enough for real time
  – takes 4 times for 4-finger slap

» Not enough levels
  – Still statistically significant

» Insensitive to partial prints
NFIQ 2.0
http://www.nist.gov/itl/iat/ig/development_nfiq_2.cfm

Modular Architecture

Flags to identify:
- Fingerness
- Completeness
- Wrong Phalanges
- Centerness
- Low/high pressure
- Alteredness

Actionable feedback

NFIQ Lite
- Fast for Mobile applications
- Use of innovative methods such as self organizing map

Mapping curves
- With NFIQ 1.0
- Calibration curve for each NFIQ 2.0 comparator participants.

Tools

`Standardized' features (ISO/IEC 29794-4)

Uses of quality assessment

Subject presentation
- Improper presentation detection
- Presentation attack detection

Acquisition device
- Hardware built-in. Quality in capture loop.
- ‘peak’ imaging capability
- No control on FTA - Hard to tweak to certain applications

Beyond scanner
- Automated (e.g., NFIQ) or visual by human
- Automated at client-side or backend
- Actionable feedback for re-capture

Operator review
- Particularly for high value images
- It is expensive
- Requires training of operators + takes time

Allows for
- Adopting threshold for specific scenario
- Monitoring Seasonal variations, atypical collection site/queue/device, etc.
Quality in large scale deployments

**National**
- DHS US-VISIT
- Low enforcement
  - FBI CJIS
- DoD

**International**
- Unique Identification Authority of India
- EU-VIS
- Law enforcement (Germany BKA)
- E-passport
Available Quality Assessment Algorithms

**NIST Open source implementation**

- **Finger**
  - NFIQ (circa 2004)
    - www.nist.gov/itl/iad/ig/nbis.cfm
    - NIST IR 7151
  - Next Generation NFIQ (NFIQ 2.0) underway
    - www.nist.gov/itl/iad/ig/development_nfiq_2.cfm
- **Latent**
- **Iris**
  - Not yet, but in near future
    - NIST Iris Image Quality (NIIQ)
    - Some methods are documented in technical literature
- **Face**
  - Not yet – no plans yet.

**Proprietary implementation**

- **Needs testing**
  - Their effectiveness in predicting performance have to evaluated.
- **Need calibration**
  - To interpret scores
  - To achieve interoperability
- **Issues with vendor lock-in**

Academia? Others?
Calibration

Calibration Curve
Quality: nfiq Dataset: poe

- Rejecting comparisons which $\min(nfiq_{sample1}, nfiq_{sample2}) = 5$
- Rejecting comparisons which $\min(nfiq_{sample1}, nfiq_{sample2}) = 4$
- Rejecting comparisons which $\min(nfiq_{sample1}, nfiq_{sample2}) = 3$
- Rejecting comparisons which $\min(nfiq_{sample1}, nfiq_{sample2}) = 2$

False non-match rate vs. False match rate graph with NIST logo.
Challenges

» Pairwise Q
» Computation time
» Different matcher – different sensitivities
» Limitation on available data, particularly data with known degree of specific impairments
WE, AT NIST
# Focus and Impact

## NIST Biometric Quality Program

### Push Towards Zero Error Biometrics

<table>
<thead>
<tr>
<th>Strengthening Science</th>
<th>Advancing metrology</th>
<th>Developing Standards</th>
<th>Developing Tool Box</th>
<th>Best Practice Guidance</th>
<th>Enumerative Bibliography</th>
<th>Coordination+Collaborations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failure Analysis</td>
<td>Performance Evaluation</td>
<td>Requirements Specifications</td>
<td>Open source Public domain</td>
<td>Instructional + Guidance</td>
<td>Technical Literature</td>
<td>Workshops, Conferences Grants (WVU, NYU Poly)</td>
</tr>
<tr>
<td>Identifying the likely causes of recognition error, quantifying their effect and ways to mitigate them.</td>
<td>Quantitative means of assessing performance of quality assessment algorithms (IREX II IQCE)</td>
<td>On image properties affecting performance, and on capture device</td>
<td>Reference implementation of quality assessment algorithm, iris segmentation</td>
<td>Materials for quality score summarization + Best capture practice + example images of various quality</td>
<td>Reports, white papers, publications relevant to biometric quality and iris image quality in particular</td>
<td></td>
</tr>
</tbody>
</table>

|----------|------------|----------|----------|--------|---------|------|----------------------------------------|
Contact Info:
Elham Tabassi
tabassi AT nist DOT gov
301 975 5292