Thoughts on Fingerprint Image Quality and Its Evaluation

NIST Biometric Quality Workshop II
November 7-8, 2007

Masanori Hara
Recap from NEC’s Presentation at Previous Workshop (2006)

- Positioning quality: a key factor to guarantee common area and matching accuracy
- Pattern area: good positioning criteria
- Quality to predict accuracy: matcher dependent (algorithm dependent)
- NEC quality metrics: better accordance with NEC matchers than NFIQ
Contents

1. Quality Concepts
2. Requirements for Quality
3. Pure Quality and Predictive Quality
4. Image Enhancement
5. Fingerprint Properness Analysis
6. Factors to Degrade Quality
7. Quality Metrics Evaluation
8. Conclusion and Suggestion

Appendix: Improvement on NEC’s Predictive Quality Metrics
1. Quality Concepts – Ideal Quality

What is ideal quality?

1) good ridge quality
   - dynamic range sufficiently wide
   - uniformity evenly distributed density
   - linearity gray mid scale reserved
   - no saturation (white & black)
   - no significant smudge or blur
   - sufficient ridge/valley separation

2) no problematic background noise
   - no Leftover fingerprint or stripe pattern
   - background lighter than ridge (foreground)

3) sufficient size
   - excellent for slap/flat matching
   - good for latent cognizant

4) good positioning and orientation
   - pattern area included & fingertip up

5) no significant distortion

Ideal quality  Strong image enhancement NOT required
1. Quality Concepts – Poor Quality Samples

Factors related to ridge quality and background noise

a) narrow dynamic range
b) uneven density
c) white saturation
d) leftover fingerprint
e) problematic stripe pattern

Note: Fingerprint samples are from NIST DB#27 and FVC Data Base.
2. Requirements for Quality

“Requirements for quality” depend on:
- target filed – **law enforcement (LE)** or **non LE (NLE)**
- operational requirement - **automatic** or **manual intervening**
- image type (flat, slap, rolled, latent)

two major operational categories:

**a) fully automatic matching (NLE/LE)**
- a1) Positive ID (cooperative & unsupervised - NLE)
- a2) Negative ID (uncooperative & supervised - NLE)
- a3) Automatic latent (rolled/latent - LE)

**b) manual intervening operation (LE)**
- b1) for latent cognizant rolled print
- b2) for latent print

*Note: a1) is not discussed here.*
2. Requirements for Quality

Criteria for quality

a) criterion for acceptance/rejection at capture
b) criterion for enrollment or registration
c) criterion for special search (e.g. latent)
d) criterion to predict “matching accuracy”

1) real-time processing required for criterion a)
2) quality metrics specific to capture device is effective for real-time processing
   e.g. positioning & orientation do NOT have to be checked for identification slap (4-slap) capturing

Note: Only offline processing for static image is discussed here.
3. Pure Quality and Predictive Quality

*pure quality* – intrinsic quality of image itself
matcher independent
also Independent of operational needs

*predictive quality* - quality for predicting accuracy
matcher dependent

a) predictive quality for auto (PQ_A)
for automatic operation (NLE, automatic latent)
*compatibility with human examiners NOT required*
*OK to use with non-minutia-based matching*

b) predictive quality for manual (PQ_M)
for manual intervening operation (LE, manual latent)
*need to consider compatibility with human examiner’s minutia definition*
4. Image Enhancement - Quality Metrics

- Question -

Image enhancement – OK for quality metrics?

1) contrast enhancement
   - contrast stretch (global & local)
   - histogram equalization (global & local)
   - sharpening, etc.
   **strong tools to cope with “narrow dynamic range” and “uneven density”, etc.**

2) ridge enhancement
   - filtering – contextual, Fourier, Gabor, Wavelet, etc.
   (*) based on ridge direction & pitch
   - pore & incipient ridge removal, etc.
   **useful to cope with insufficient ridge/valley separation**
4. Image Enhancement - Quality Metrics

- Question -

*Image enhancement - OK for quality metrics?*

No, at least for pure quality!

Image enhancement has side effects such as:

1) increasing background noise
2) removing gray intermediate (mid) scale
3) creating false (ghost) ridges
4) removing true ridges
5) creating false minutiae and missing true minutiae
4. Image Enhancement – Contrast Enhance

Which image is of better quality?

- Image A
  - Dynamic range: narrower
  - Gray mid scale: equivalent
  - Background noise: less

- Image B
  - Dynamic range: wider
  - Gray mid scale: equivalent
  - Background noise: more

Transformed NFIQ employed for easier comparison:

<table>
<thead>
<tr>
<th>NFIQ (original):</th>
<th>poorer</th>
<th>better</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFIQ (transformed):</td>
<td>0 25 50 75 100</td>
<td></td>
</tr>
<tr>
<td>NEC quality metrics:</td>
<td>0 -------------------100</td>
<td></td>
</tr>
</tbody>
</table>

Image B: contrast-enhanced image of Image A using local contrast stretch
4. Image Enhancement – Contrast Enhance

Which image is of better quality?

Image A

dynamic range
narrow < wide
gray mid scale
some > no

NFIQ=100; NEC=100

Image B: contrast-enhanced image of Image A using sharpening and contrast stretch

no difference between pore and valley!

NFIQ=75
NEC=100
4. Image Enhancement – Contrast Enhance

Which image is of better quality?

- **Image A**: contrast-enhanced image of Image A using sharpening and contrast stretch
- **Image B**: contrast-enhanced image of Image A using sharpening and contrast stretch

No difference between incipient ridge and true ridge!
4. Image Enhancement – Contrast Enhance

Which image is of better quality?

A

<table>
<thead>
<tr>
<th>dynamic range</th>
<th>narrow &lt; wide</th>
</tr>
</thead>
<tbody>
<tr>
<td>gray mid scale</td>
<td>some &gt; little</td>
</tr>
<tr>
<td>background noise</td>
<td>more &lt; less</td>
</tr>
<tr>
<td>white saturation</td>
<td>no &gt; yes</td>
</tr>
</tbody>
</table>

NFIQ=75; NEC=39

B

NFIQ=75; NEC=30

Image B: contrast-enhanced image of Image A using *light density removing* and contrast stretch
4. Image Enhancement – Ridge Enhance

Which image is of better quality?

Image A
- NFIQ=50; NEC=61
- dynamic range: narrow < wide
- gray mid scale: yes > not reliable
- fidelity: Yes > No

Image B
- NFIQ=75; NEC=80
- Ridge-enhanced image does NOT ALWAYYS represent the original image.

Image B: ridge-enhanced image of Image A using contextual filtering
4. Image Enhancement – Ridge Enhance

Which image is of better quality?

Image A: ridge-enhanced image using fixed-width pitch data (false ridges)
Image B: ridge-enhanced image using fixed-width pitch data (false ridges)
Image C: ridge enhanced image using variable-width pitch data (locally estimated)

Ridge-enhanced image does NOT ALWAYS represent the original image.
4. Image Enhancement – Predictive Quality

*Image enhancement - OK for predictive quality?*

a) predictive quality for auto (PQ_A)

*no problem to apply any image enhancements*
reasonable to use equivalent method to matcher
(1) false minutia per examiners’ definition be useful as long as such “feature” is consistent
(2) consistently miss-extracting true (but unstable) minutiae be more favorite than inconsistent extraction

b) predictive quality for manual (PQ_M)

*limited use suggested*
strong filtering in ridge enhancement tends to create false minutiae or to remove true minutiae
4. Image Enhancement – Predictive Quality

**PQ_A: strong filtering - robust to low quality image**

Ignoring *unstable minutiae* increases *consistency of minutia* on auto process.

- **A**: NFIQ=100; NEC=85
- **B**: NFIQ=100; NEC=74
- **C**: NFIQ=75; NEC=66

Nearly ideal minutiae (77: number of minutiae)

*strong filter (55)*

*strong filter (56)*
4. Image Enhancement – Predictive Quality

**PQ_A: strong filtering - not desired for manual latent**

Both unstable minutiae (e.g., crossover in red circle) and stable minutiae (in yellow circle) are important as latent cognizant features for manual latent.

- Weak filtering in ridge enhancement:
  - Good for manual latent
- Strong filtering in ridge enhancement:
  - Not desired for manual latent

Unstable minutiae missing.
4. Image Enhancement – Predictive Quality

**PQ_A: strong filtering – why robust?**

Strong filtering in ridge enhancement creates stripe patterns even though there is no real ridge information on the input image.

This process also tends to create pseudo ridges (maybe false ridges) or to remove true ridges.
4. Image Enhancement – Latent

pure quality for latent print    not practical to be evaluated
predictive quality for manual latent    dependent on human examiner

\textbf{PQ\_A: noise reduction - necessary for auto latent}

Some latent prints have severe background noise. Quality metrics as well as matching expectation for those latent prints depend on \textit{noise reduction performance}
4. Image Enhancement – Latent

**PQ_A: filtering - necessary for auto latent**

Filtering is robust to noise. It tends to create false ridge even from non-fingerprint patterns such as smudge.
5. Fingerprint Properness Analysis

Fingerprint properness and of matching feature sufficiency, etc. are important for predictive quality metrics.

1) fingerprint pattern?

- Non fingerprint
  - NFIQ=75
  - NEC=30

2) Matching feature sufficient?

- No minutia pattern
  - NFIQ=75
  - NEC=0

3) Rotated?

Accepting rotated images will unnecessary increase matching cost and unnecessary decrease matching accuracy.

- NFIQ=100
  - NEC=33

- NFIQ=100
  - NEC=85
5. Fingerprint Properness Analysis

4) distortion?

**difficult to evaluate from static image!**

<effective countermeasures>
1) identification slap (4-slap) to prevent “intentional” distortion
2) matcher algorithm improvement but, increase in cost involved

Note: As for positioning and size issues, please see our presentation at 2006 Workshop
6. Factors to Degrade Quality

<table>
<thead>
<tr>
<th>Factors to Degrade Quality</th>
<th>Possible Causes</th>
<th>Effective Countermeasures</th>
<th>Recapture Useful?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) poor ridge quality</td>
<td>device performance: 40%</td>
<td>better device supervision</td>
<td>to some extent</td>
</tr>
<tr>
<td></td>
<td>capture operation: 30%</td>
<td>periodical replacement</td>
<td></td>
</tr>
<tr>
<td></td>
<td>device problem: 20%</td>
<td>nothing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>nature of skin: 10%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2) background noise</td>
<td>device performance: 50%</td>
<td>better device maintenance (clean up)</td>
<td>to some extent</td>
</tr>
<tr>
<td></td>
<td>capture operation: 50%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3) insufficient size</td>
<td>device performance: 20%</td>
<td>large area capture device</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>capture operation: 80%</td>
<td>4-slap (3/2-slap) capture supervision</td>
<td></td>
</tr>
<tr>
<td>4) Improper positioning or orientation</td>
<td>device performance: 20%</td>
<td>4-slap (3/2-slap) capture</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>capture operation: 80%</td>
<td>large area capture device supervision</td>
<td></td>
</tr>
<tr>
<td>5) distortion</td>
<td>device performance: 50%</td>
<td>4-slap (3/2-slap) capture</td>
<td>limited</td>
</tr>
<tr>
<td></td>
<td>capture operation: 50%</td>
<td>supervision</td>
<td></td>
</tr>
</tbody>
</table>
6. Factors to Degrade Quality

effective countermeasures

1) better device most effective countermeasure
identification slap capture (4-slap)
- to consistently capture pattern area
- to solve rotation problem (fingertip up)
- to avoid distortion
- to avoid wrong finger capture

(*) 3 or 2-slap capturing is also effective

2) better supervision and capture operation
- to reduce background noise (by platen clean up, etc.)
- to capture sufficient size, etc.

Ref. T. Hopper; Identification Flats (NIST Fingerprint Standard Workshop)
7. Quality Metrics Evaluation

evaluation method - different per type of quality metrics

- pure quality metrics evaluation
  
  no straightforward method  not good for contest

  recommendation
  specific criteria be evaluated by specific algorithms

- predictive quality metrics evaluation
  
  matcher dependent  tied up with matcher

  recommendation for contest
  RRG(99.9): Rejection Rate to Guarantee 99.9%
  quality and matcher integrated evaluation
7. Quality Metrics Evaluation

pure quality metrics evaluation

Which image is of better quality?
A) good ridge quality but fingertip only
B) poor ridge quality but pattern area exits

difficult to define overall quality rating!

Can you rank these images?

NFIQ=75; NEC=28
NFIQ=50; NEC=26

FVC2002DB3 22_6
FVC2002DB3 22_2
7. Quality Metrics Evaluation

pure quality metrics evaluation

specific criteria be evaluated by specific algorithms

1) ridge quality  specific check tool
    simple method suggested  NIST (or public domain) open source

2) background noise  specific check tool
    simple method suggested  NIST (or public domain) open source

3) size  specific check tool
    simple method suggested  NIST (or public domain) open source

4) positioning and orientation  specific check tool
    good candidate for contest
    reference (correct) data needed (manual coded) for contest

5) distortion  difficult to check
    difficult to evaluate from static image
7. Quality Metrics Evaluation

predictive quality metrics evaluation

1) PQ_A evaluation is relatively Simple. This is discussed here.

2) However, PQ_M evaluation is difficult. Method for pure quality evaluation is also practical for PQ_M evaluation.
7. Quality Metrics Evaluation

recommended criteria for PQ_A metrics evaluation

RRG(X): Rejection Rate to Guarantee X% Accuracy

Given:

1) A set of fingerprint images
2) Its accuracy is less than X% (e.g. 99.9%)

Question:

How much proportion of the poorest quality images need to be rejected in order to guarantee X% (e.g. 99.9%) accuracy?

RRG(X) – straightforward criteria to evaluate the predictive quality (PQ_A) metrics and matching performance at the same time
7. Quality Metrics Evaluation

RRG(99.9) - recommended evaluation method

1) contestants provide three programs
   a) a quality program to produce quality metrics (e.g. Q: 0-100)
   b) a feature extraction program to produce templates
   c) a matching program to produce score (e.g. 0-9999)

2) NIST conducts test at the NIST facility
   a) produces quality metrics for search and file \( Q_{\text{search}}, Q_{\text{file}} \)
   b) TAR (or first rank hit) considered for simplicity
   c) determine RRG(99.9) as follows
      reject a mate if \( Q_{\text{search}} < Q_{\text{th}} \) or \( Q_{\text{file}} < Q_{\text{th}} \); \( Q_{\text{th}} : Q_{\text{threshold}} \)
      calculate \( R(Q_{\text{th}}) \) - percentage the rejection (a function of Q)
      find RRG(99.9) – minimum R to achieve 99.9% accuracy

Note: 99.9% (instead of 100%) is recommended as target accuracy in order to avoid undesired side effect from the exceptional data
7. Quality Metrics Evaluation

Sample of RRG(X): R(Q) vs. TAR

straightforward method to compare different algorithms
8. Conclusion and Suggestion

1) image enhancement for pure quality
   - shall NOT be used or limited to moderate contrast enhancement

2) image enhancement for predictive quality
   - for PQ_A (automatic)   no restriction
   - for PQ_M (manual)     limited use suggested

3) evaluation for pure quality
   - specific algorithm be developed by NIST (or public domain)
   - not appropriate for contest

4) evaluation for predictive quality (PQ_A)
   - RRG(99.9) criteria suggested
   - appropriate for contest with proprietary matcher

5) practical solution for negative identification system
   - identification slap capture (4-slap capture)
Appendix: Improvement on NEC’s Predictive Quality Metrics (1/13)

NEC quality metrics  PQ_A

- Rated on a 0-100 scale, where 0 is the lowest quality and 100 is the highest quality
- Nonlinear combination of four independent indices
  - ridge quality with its area size
  - high-confidence minutiae count
  - positioning quality for common area
  - distortion tolerance

Note: Appendix is prepared by Amane Yoshida from his research.
### Accuracy Improvement

<table>
<thead>
<tr>
<th>Speed</th>
<th>Match</th>
<th>FE</th>
<th>DB1</th>
<th>DB2</th>
<th>DB3</th>
<th>DB4</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-equiv.</td>
<td>Slow</td>
<td></td>
<td>99.73 (+0.09)</td>
<td>99.93 (+0.18)</td>
<td>99.07 (+0.69)</td>
<td>99.80 (+1.09)</td>
</tr>
<tr>
<td>SDK H2</td>
<td>H-equiv.</td>
<td>H-equiv.</td>
<td>99.02</td>
<td>99.68</td>
<td>92.13</td>
<td>96.36</td>
</tr>
<tr>
<td>SDK H</td>
<td>See NISTIR7151</td>
<td></td>
<td>99.02</td>
<td>99.68</td>
<td>92.13</td>
<td>96.36</td>
</tr>
</tbody>
</table>
Quality Distributions

FVC2002 DB1

FVC2002 DB2

FVC2002 DB3

FVC2002 DB4
Appendix: FVC2002 (4/13)

DB1: RRG comparison over varying FAR
DB2: RRG comparison over varying FAR

![Graphs showing DB2 RRG comparison over varying FAR](image)

---

(c) NEC Corporation 2007
Biometric Quality Workshop II
Empowered by Innovation

Page 39
Appendix: FVC2002 (6/13)

DB3: RRG comparison over varying FAR

![Graphs showing DB3 RRG comparison over varying FAR](image_url)
DB4: RRG comparison over varying FAR

![Graphs showing RRG comparison for DB4 at different FAR](image-url)
### Accuracy Improvement

<table>
<thead>
<tr>
<th>Speed</th>
<th>Match</th>
<th>FE</th>
<th>TAR at FAR=0.01% [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>DB1</td>
</tr>
<tr>
<td>SDK H4 (2007)</td>
<td>H-equiv.</td>
<td>Slow</td>
<td>96.68 (+0.93)</td>
</tr>
<tr>
<td>SDK H3 (2006)</td>
<td>H-equiv.</td>
<td>Slow</td>
<td>95.75</td>
</tr>
<tr>
<td>SDK H2</td>
<td>H-equiv.</td>
<td>H-equiv.</td>
<td>95.66</td>
</tr>
<tr>
<td>SDK H</td>
<td>See NISTIR7151</td>
<td></td>
<td>93.63</td>
</tr>
</tbody>
</table>
Quality Distributions

FVC2004 DB1

FVC2004 DB2

FVC2004 DB3

FVC2004 DB4
DB1: RRG comparison over varying FAR

- FVC2004 DB1 (FAR=0.1%)
- FVC2004 DB1 (FAR=0.01%)
DB2: RRG comparison over varying FAR
Appendix: FVC2004 (12/13)

DB3: RRG comparison over varying FAR

![Graphs showing RRG comparison over varying FAR for DB3 from FVC2004](image-url)
Appendix: FVC2004 (13/13)

**DB4: RRG comparison over varying FAR**

![Graphs showing RRG comparison over varying FAR](image_url)