Biometric Sample Quality

The last 1% - Biometric quality assessment for error suppression

Elham Tabassi
Biometric Consortium
September 12, 2007

motivation

- Performance of biometric systems depends on the quality of the acquired input samples
  - performance degrades substantially as quality degrades
  - although only a small fraction of input data are of poor-quality, the bulk of recognition errors can be attributed to poor-quality samples

- If quality can be improved, either by sensor design, by user interface design, or by standards compliance, better performance can be realized.

- For those aspects of quality that cannot be designed-in, an ability to analyze the quality of a live sample is needed. This is useful in
  - initiating the reacquisition from a user,
  - the real-time selection of the best sample,
  - the selective invocation of different processing methods.
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.

quality is not about human perception

NFIQ=5  NFIQ=2
four types of quality

1. character, e.g. scar
2. behavior, e.g. pose
3. imaging, e.g. focus
4. environment, e.g. shadows

uses of quality to improve performance

quality directed processing

quality monitoring

conditional reacquisition
conditional reacquisition::how to set quality threshold?

set verification quality threshold based on error-vs-reject curves

quality monitoring::performance related quality summarization

quality summarization supports monitoring
- over time (to expose seasonal variation, or trends),
- for each sensor (to identify defective devices),
- at each site (to identify problem locations)
- of officials or attendants (to assess adherence to operating procedures), and
- per user basis (to identify users that consistently yield low quality samples)

In each case the quality summaries can be used to identify departures from the application specific historical norms, or design targets.
**enterprise-wide quality values**

Summarizing quality values computed across all retained samples in an enterprise into a single quality value representing the overall quality of the enterprise.

**Black or Gray Box Quality summarization:**

- \( q_1, q_2, \ldots, q_i, \ldots, q_N \)

**performance related quality summarization**

- Is not well represented by arithmetic mean of quality values
- Should be a weighted average of the native quality values
  - Weights for each quality level \( q \) \( ( \mu_q ) \) should be directly related to the error rate observed for samples of quality \( q \)
- Should be on the range \([0, 100]\)
  - As specified in ISO/IEC 19784-1 BioAPI and ISO/IEC 29794-1
- Could be the result of a biometric quality assessment calibration process
  - Conducted by the provider, or by a third party laboratory
- Should be performed across similar usage
  - E.g., quality summarization over all enrollment samples of an enterprise or per user basis in time and attendance applications
Best practice NFIQ summarization

\[ Q = 102.75 \cdot 2.75p_1 - 5.37p_2 - 14.38p_3 - 42.25p_4 - 102.75p_5 \]

- \( p_i \) is the proportion of the fingerprints with NFIQ values \( i \in \{1, 2, \ldots, 5\} \) in the enterprise.
- the weights reflect the likelihood that an observed false non-match involved a fingerprint of quality \( i \)
  - weights were estimated using the observed false non-match rates from a set of leading commercial matching algorithms computed at some fixed threshold

---

NFIQ::(fmr,fnmr) at fixed threshold
computation of weights

1. for all V matching algorithms compute false non-match rate for each L quality levels at threshold $\tau$

   \[
   \text{FNMR}(\tau, i) = \begin{cases} 
   q_i^a & \text{if } q_i^a < \tau, q_i^b < i, q_i^b > i \\
   q_i^b & \text{if } q_i^a > \infty, q_i^b > i, q_i^b = i 
   \end{cases}
   \]

   which results in the following array

   \[
   \begin{pmatrix}
   \text{FNMR}(\tau_1, 1) & \text{FNMR}(\tau_1, 2) & \cdots & \text{FNMR}(\tau_1, L) \\
   \text{FNMR}(\tau_2, 1) & \text{FNMR}(\tau_2, 2) & \cdots & \text{FNMR}(\tau_2, L) \\
   \cdots & \cdots & \cdots & \cdots \\
   \text{FNMR}(\tau_L, 1) & \text{FNMR}(\tau_L, 2) & \cdots & \text{FNMR}(\tau_L, L)
   \end{pmatrix}
   \]

2. $u_i, i = 1, \ldots, L$ is normalized false non-match

   \[
   q_i = \frac{\text{FNMR}(\tau, i)}{\sum_{i=1}^{L} \text{FNMR}(\tau, i)}
   \]

3. Map (e.g. linear mapping of $[u_1, \ldots, u_L]$ to $[0, \ldots, 100]$, for example NFIQ summarization is given by:

   \[
   Q = \frac{100q}{q_1} + \frac{100q_2}{q_2} + \cdots + \frac{100q_L}{q_L}
   \]
Dedicated weights should be computed in verification applications, where

- a specific set of one or more matching algorithms are known and available, or

- operating threshold is known and different from weights for Best Practice NFIQ summarization.

---

**some examples :: quality monitoring**

<table>
<thead>
<tr>
<th></th>
<th>#samples</th>
<th>NFIQ 1</th>
<th>NFIQ 2</th>
<th>NFIQ 3</th>
<th>NFIQ 4</th>
<th>NFIQ 5</th>
<th>NFIQ summary</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>across airports</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>795</td>
<td>197</td>
<td>322</td>
<td>240</td>
<td>25</td>
<td>11</td>
<td>92.80</td>
</tr>
<tr>
<td>L</td>
<td>929</td>
<td>283</td>
<td>358</td>
<td>255</td>
<td>18</td>
<td>15</td>
<td>93.42</td>
</tr>
<tr>
<td>M</td>
<td>1346</td>
<td>473</td>
<td>536</td>
<td>302</td>
<td>22</td>
<td>13</td>
<td>94.74</td>
</tr>
<tr>
<td><strong>per user</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>4</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>52.24</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>16</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>99.48</td>
</tr>
</tbody>
</table>
### More Examples :: Meta Data Analysis

<table>
<thead>
<tr>
<th>Age</th>
<th>Younger than 30</th>
<th>Between 30 and 65</th>
<th>Older than 65</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Samples</td>
<td>78685</td>
<td>98312</td>
<td>6452</td>
</tr>
<tr>
<td>NFIQ Summary</td>
<td>98.20</td>
<td>95.60</td>
<td>85.40</td>
</tr>
<tr>
<td>95% Confidence Interval</td>
<td>(98.14,98.25)</td>
<td>(95.48,95.70)</td>
<td>(84.82,85.98)</td>
</tr>
</tbody>
</table>

Data: poe-bva :: 183449 right index

<table>
<thead>
<tr>
<th>Gender</th>
<th>All</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFIQ Summary</td>
<td>94.93</td>
<td>95.98</td>
<td>93.80</td>
</tr>
<tr>
<td>95% Confidence Interval</td>
<td>(95.86,96.05)</td>
<td>(93.67,93.93)</td>
<td></td>
</tr>
</tbody>
</table>

Data: poe-bva :: 183449 left index

---

### Summary

- Quality measurement has vital roles in improving biometric system accuracy and efficiency
  - During the capture process (as a control-loop variable to initiate reacquisition)
  - In enterprise wide quality-assurance surveying
  - In invocation of quality-directed processing of samples

- Quality threshold
  - FNMR, FMR, and FTA should be considered

- Quality summarization is operationally useful
  - Quality summary statistic should be indicative of performance
    - For verification, it should serve as a measure of overall expected false non-match rate
quality by design
- To what extend does good design guarantee good quality?
- Goal: drive acquisition errors to zero

quality calibration
- Performance based mapping from vendors’ quality score to standard range of [0,100]
- Calibration to a specific matching algorithm or calibration for general use
- Goal: quality score interpretation and interoperability by relating quality scores to performance

quality evaluation
- How to assess performance of biometric quality measures?
- Goal: biometric quality evaluation

quality reference data
- Would quality annotated corpora be useful in quality interoperability or quality evaluation?
- If useful and needed, how to build a quality reference data set?

quality standardization
- Should SC37 consider adding Best Capture Practice to existing data format standards?
- To what extend should conformance to data format standards be indicative of quality?
- Is standardizing performance testing of quality measurement algorithms needed?
Thank You

The quality summarization report is on line
www.itl.nist.gov/iad/894.03/quality/reports/enterprise.pdf

NIST Quality Program
www.itl.nist.gov/iad/894.03/quality/index.html

Feedback is welcomed.

tabassi@nist.gov - BC September, 12 2007