Independent Testing of Iris Recognition Technology (ITIRT)

Michael Thieme
Director of Special Projects
International Biometric Group

Biometric Consortium Symposium
19 September 2005
About International Biometric Group

- Independent biometric integration, consulting, and research & testing company
  - Design, integrate, deploy biometric-enabled applications
  - Test biometric devices and systems
  - Consult on biometric projects and technologies for government, commercial organizations
  - Publish biometric technology and market research

- Areas of expertise
  - State of the art in biometric modalities and products
  - Standards and interoperability
  - Privacy, policy, and legal issues
  - User interface, integration of biometrics into workflow
  - Multiple biometrics and fusion
Iris Recognition Performance Evaluation

- Low FMR shown in previous evaluations, including those executed by:
  - Dr. John Daugman, *various reports*
  - Iridian, *Cross-Comparison Testing*
  - International Biometric Group, *Comparative Biometric Testing*

- Limitations of previous iris testing
  - Vendor-executed tests focus largely on impostor comparisons (i.e. no FNMR, FRR)
  - Few tests address acquisition rates or level of effort (FTA, FTE)
  - Independent tests are small-scale (CESG-NPL, IBG)
  - Testing limited by access to databases and vendor technologies
ITIRT Overview
Background

- IBG contracted by US Department of Homeland Security (Transportation Security Administration) to execute ITIRT
  - A test of the accuracy, usability, and interoperability of iris recognition systems used in access control / civil ID
- Test executed 29 July 2004 – 2 May 2005
  - Test platform buildout, integration (3 months)
  - Subject recruitment and sample collection (2 months)
  - Matching and data analysis (2 months)
  - Report generation (2 months)
- ITIRT Report released by DHS 1 June 2005
  - www.biometricgroup.com/ITIRT
- Funding / executing org had no stake in performance of system(s) that would impact test methodology, execution
  - Vendors involved to validate design, code, methodology
Objectives

- Acquire samples through different vendors’ systems, match through a single algorithm
- Enroll in real time, match offline
- Compare samples within and across each system
  - *Intra-system* establishes real-world accuracy baseline
  - *Cross-system* gathers data on interoperable accuracy
- Execute comparisons based on enrollment and recognition data
- Assess transactional usability, capture rates, effort level
  - Identify tradeoffs between accuracy, ease of use
- Automate data collection and application logic
  - Reduce operator error and variability
- Evaluate test subject-specific trends
  - E.g. susceptibility to genuine or impostor errors
Scale

- Test subjects
  - 1224

- Visits
  - 1682
  - 766 1-visit subjects, 458 2-visit subjects

- Templates (i.e. IrisCodes) and Samples
  - Up to 16 IrisCodes acquired per test subject during enrollment
  - Up to 54 iris samples acquired per test subject during recognition
  - ~110,000 samples and IrisCodes collected in course of test

- 1:1 comparisons
  - Genuine comparisons: 928,916
  - Impostor comparisons: 1,991,835,788
Days Elapsed between Visits 1 and 2
Iris Recognition
Hardware and Software
Acquisition Devices / Systems

LG IrisAccess 3000

- 6 acquisition systems provided by LG, OKI, Panasonic
  - Separate devices used for enrollment and recognition
  - Networked to dedicated workstations running capture application

- Device selection based on market presence, use in access control / ID systems
Capture, Enrollment, and Matching Software

- IBG used vendor-provided software tools to build a custom test environment
- Iridian KnoWho OEM SDK
  - Used to build custom enrollment and matching application
  - Generates full range of match scores, a test requirement
  - Processes PrivateID-formatted IrisCodes, samples
- Device-specific APIs, SDKs
  - Used to build capture and enrollment applications
  - Panasonic PrivateID 2.3; OKI PrivateID 2.2
    - OKI and Panasonic HW also available in non-PrivateID versions; performance may vary, in particular transaction times
  - LG IrisAccess 3000 SDK v3.0
    - LG 3000 device does not generate PrivateID-format IrisCodes
    - Iridian provided utility to convert LG bitmaps to PrivateID format
Sample Collection and Comparison Processes
# Test Subject-Device Interaction

<table>
<thead>
<tr>
<th>Model</th>
<th>Presentation Method</th>
<th>Acquisition Method</th>
<th>Imager Adjustable</th>
<th>Subject Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>LG IrisAccess 3000 EOU, ROU</td>
<td>L/R irises presented separately</td>
<td>L/R irises acquired separately</td>
<td>Yes</td>
<td>Spoken / Visual</td>
</tr>
<tr>
<td>OKI IRISPASS-WG</td>
<td>L/R irises presented simultaneously</td>
<td>L/R irises acquired separately</td>
<td>No</td>
<td>Tone / Visual</td>
</tr>
<tr>
<td>Panasonic BM-ET300</td>
<td>L/R irises presented simultaneously</td>
<td>L/R irises acquired simultaneously</td>
<td>Yes</td>
<td>Spoken / Visual</td>
</tr>
</tbody>
</table>

- **LG**: rapid, intuitive capture, but a single-eye imager
- **OKI**: usable with almost no training, time-consuming
- **Panasonic**: rapid dual-eye capture; can be difficult to use
Enrollment / Recognition Sample Collection

• Enrollment
  – Test subjects (seated) given real-time guidance
  – Application cycled until irises enrolled or timeout declared
    • Acquisition of 2-4 sufficient-quality samples required
    • Device may acquire 6 samples to find 3 of sufficient quality
  – Systems declared FTE in real time
    • If < 2 irides enrolled, 1 full additional transaction executed
  – Every test subject was enrolled twice on each system
  – Up to 16 IrisCodes enrolled / subject: 4 OKI, 4 Panasonic, 8 LG
  – Enroll on all, then present recognition samples on all

• Recognition Sample Collection
  – Test subjects (standing) presented samples
    • Instructions not provided except in cases of egregious misuse
  – Recognition samples evaluated in real time, matched offline
    • If FTA, no additional attempts permitted
    • Acquisition based on LG, Panasonic, OKI criteria
  – Multiple transactions: up to 54 samples collected / subject
Why Collect Multiple Samples per Subject?

- For many systems, genuine accuracy (FNMR) poses greater challenges than impostor accuracy (FMR)
  - Much easier to generate impostor scores than genuine scores
  - Assuming two samples per subject and 100 subjects, a test can generate ~20,000 impostor scores but only ~100 genuine scores
- Multiple transactions increase # of genuine comparisons
  - Enables test cases that otherwise cannot be executed
    - E.g. transaction-level and attempt-level performance
    - Same-day vs. different-day performance
  - Can measure impact of increased familiarity with devices
- Limitation
  - Statistically not the same as testing on a very large population
## Comparison Types

<table>
<thead>
<tr>
<th>Comparison Type</th>
<th>Probe Dataset</th>
<th>Gallery Dataset</th>
<th>Excluded Comparisons</th>
</tr>
</thead>
</table>
| **Enrollment Comparison** | 24,627 Enrollment IrisCodes | 24,627 Enrollment IrisCodes | - Same IrisCode  
- Same-subject left vs. right iris |
| **Recognition Comparison** | 87,275 Recognition Samples | 24,627 Enrollment IrisCodes | Same-subject left vs. right iris |

**Enrollment Comparison**: Enrollment IrisCodes vs. Enrollment IrisCodes  
**Recognition Comparison**: recognition samples vs. Enrollment IrisCodes  
**Intra-device**: acquired through *same* vendor devices  
**Cross-device**: acquired through *different* vendor devices  
**Intra-visit**: acquired on the *same* day  
**Cross-visit**: acquired *days or weeks* apart
Results
## Failure to Enroll Rates

<table>
<thead>
<tr>
<th></th>
<th>LG IrisAccess 3000</th>
<th>OKI IRISPASS-WG</th>
<th>Panasonic BM-ET300</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Transactions</strong></td>
<td>3364</td>
<td>3364</td>
<td>3364</td>
</tr>
<tr>
<td><strong>Zero Irides Enrolled (FTE)</strong></td>
<td>41</td>
<td>1.61%</td>
<td>85</td>
</tr>
<tr>
<td><strong>Either Left or Right Iris Enrolled</strong></td>
<td>244</td>
<td>8.50%</td>
<td>262</td>
</tr>
<tr>
<td><strong>Both Irides Enrolled</strong></td>
<td>3079</td>
<td>89.89%</td>
<td>3017</td>
</tr>
</tbody>
</table>

**FTE**: % of enrollment transactions in which zero irides were enrolled
## Transaction-Level Failure to Acquire Rates

<table>
<thead>
<tr>
<th></th>
<th>LG IrisAccess 3000</th>
<th>OKI IRISPASS-WG</th>
<th>Panasonic BM-ET300</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Transactions</strong></td>
<td>5040</td>
<td>5031</td>
<td>5043</td>
</tr>
<tr>
<td><strong>Zero Iris Samples Acquired</strong></td>
<td>35  0.69%</td>
<td>16  0.32%</td>
<td>21  0.42%</td>
</tr>
<tr>
<td>≥ 1 Left or ≥ 1 Right Iris Samples Acquired</td>
<td>89  1.77%</td>
<td>46  0.91%</td>
<td>55  1.09%</td>
</tr>
<tr>
<td>≥ 1 Left and ≥ 1 Right Iris Samples Acquired</td>
<td>4916  97.54%</td>
<td>4969  98.77%</td>
<td>4967  98.49%</td>
</tr>
</tbody>
</table>

**T–FTA:** % of recognition transactions with zero sufficient-quality samples acquired in 6 attempts (3 left, 3 right)
Error Rates at 0.33 HD

<table>
<thead>
<tr>
<th></th>
<th>FNMR</th>
<th>FMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>LG-LG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OKI-OKI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAN-PAN</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- 0.33 HD: threshold used for 1:1 matching
- Of note: type of comparison (enrollment or recognition) a larger factor than time
Enrollment Comparison DETs

- LG-LG: 0.00058 FMR, 0.964 FNMR
- LG-OKI: 0.00014 FMR, 1.117 FNMR
- PAN-PAN: 0.00013 FMR, 0.759 FNMR
- OKI-OKI EER: 0.22738 FMR, 0.246 FNMR
- PAN-PAN EER: 0.21849 FMR, 0.235 FNMR

False Match Rate (FMR) vs. False Non-Match Rate (FNMR)
Recognition Comparison DETs

![Graph showing DET comparison for different biometric systems including OKI-OKI, PAN-PAN, LG-LG, and others. The graph plots False Non-Match Rate (FNMR) against False Match Rate (FMR) with various EER (Equal Error Rate) values indicated for each system.]
### Cross-Visit Enrollment at Declining HDs

<table>
<thead>
<tr>
<th></th>
<th>Cross-Visit Enrollment Comparisons</th>
<th>1:1</th>
<th>1:100</th>
<th>1:10,000</th>
<th>1:1,000,000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.330 HD</td>
<td>0.310 HD</td>
<td>0.290 HD</td>
<td>0.270 HD</td>
</tr>
<tr>
<td>LG-LG</td>
<td>FNMR</td>
<td>1.636%</td>
<td>2.104%</td>
<td>2.821%</td>
<td>3.880%</td>
</tr>
<tr>
<td></td>
<td>FMR</td>
<td>0.00058%</td>
<td>0.00003%</td>
<td>0.0000%</td>
<td>0.0000%</td>
</tr>
<tr>
<td>OKI-OKI</td>
<td>FNMR</td>
<td>1.980%</td>
<td>2.833%</td>
<td>3.929%</td>
<td>6.092%</td>
</tr>
<tr>
<td></td>
<td>FMR</td>
<td>0.00001%</td>
<td>0.0000%</td>
<td>0.0000%</td>
<td>0.0000%</td>
</tr>
<tr>
<td>PAN-PAN</td>
<td>FNMR</td>
<td>1.422%</td>
<td>2.031%</td>
<td>3.114%</td>
<td>5.146%</td>
</tr>
<tr>
<td></td>
<td>FMR</td>
<td>0.00031%</td>
<td>0.00001%</td>
<td>0.0000%</td>
<td>0.0000%</td>
</tr>
<tr>
<td>OKI-PAN</td>
<td>FNMR</td>
<td>4.031%</td>
<td>6.517%</td>
<td>9.405%</td>
<td>13.369%</td>
</tr>
<tr>
<td></td>
<td>FMR</td>
<td>0.00015%</td>
<td>0.0000%</td>
<td>0.0000%</td>
<td>0.0000%</td>
</tr>
<tr>
<td>PAN-OKI</td>
<td>FNMR</td>
<td>5.382%</td>
<td>7.916%</td>
<td>10.920%</td>
<td>14.393%</td>
</tr>
<tr>
<td></td>
<td>FMR</td>
<td>0.00022%</td>
<td>0.0000%</td>
<td>0.0000%</td>
<td>0.0000%</td>
</tr>
</tbody>
</table>

- FMR disappears; FNMR increases steadily
- 0.01 reduction in HD = order of magnitude larger DB
Non-Same Iris Pair with HD < 0.33
Mean Genuine HD Distribution

Based on ~170 comparisons per test subject
Mean Impostor HD Distribution

Based on ~450,000 comparisons per test subject
Conclusions

- 1:1 FMR ranged from ~1:100k to ~1:1m
- Certain devices more resistant to FMR than others
- Both FNMR and capture failures are an issue
  - More so than for other modalities, static databases only address a fraction of the overall performance problem
- Samples interoperable across devices, but genuine errors increase relative to same-device performance
- Certain test subjects have high genuine error rates; all subjects have consistently low impostor error rates
Areas for Further Study

- **Iris quality**
  - What accounted for genuine error rates with enrollment IrisCodes?
  - What is the relationship between iris quality and FMR / FNMR?

- **Test subject-specific trends**
  - Why are certain people outliers – behaviour or physiology?

- **Eye colour and ethnicity**
  - Does any data suggest that eye colour or ethnicity are associated with better / worse performance?

- **ITIRT dataset matched through alternative algorithms**

- **Multimodal testing with face, fingerprint data to examine multi-modal correlation**
Contact Information

mthieme@biometricgroup.com