How AFIS Selection Was Performed for IAFIS: History and Lessons Learned

April 6, 2006
George Kiebuzinski
Pre IAFIS Concerns

- IDAS ten-print matching was poor (TAR of about 90%, system FAR of about $10^{-2}$ for database of 30M versus current system FAR of better than $10^{-9}$ for a database of 42M)
- Fingerprint pattern classification was very manually intensive
- Latent matching performance was very poor & very slow
- Database and workload sizes greater than any existing AFIS
- Database updates were very inefficient
  - Data was binned
  - Binning required periodic reloads of the entire database
- The databases in existing AFISs were overwhelmed, not able to scale (to differing extents), processing was frequently unbalanced
Background

• AFIS source selection started c. 1994
• Need to replace IDAS
  – Heavy reliance on soft biometrics
  – Minimal latent search capability
  – Manual pattern and sub-pattern classification (Extended Henry set)
• Requirements published early; industry and LEAs comments incorporated into RFP
• Two stage procurement RFP
  – Basic Demonstration Model (BDM) - three vendors to be funded to demonstrate critical technology and risk mitigation
  – Final down-select – one vendor to build complete system
Maximizing competition

- The first stage – BDM required written proposals and orals
- Five teams competed:
  - TRW/Cogent / note that TRW/Cogent had recently won the UK’s NAFIS contract
  - Martin Marietta Data Systems/SAGEM Morpho
  - UNISYS/NEC
  - Westinghouse/Printrak
  - Calspan [note that Calspan had built a significant portion of the previous FBI fingerprint identification system]
- BDM winners were:
  - TRW /Cogent
  - Martin Marietta/SAGEM Morpho/Calspan (later Lockheed Martin team)
  - UNISYS/NEC
- In all cases, the system integrators were the dominant partners
- All BDM participants were required to revise their proposals incorporating BDM results
The FBI funded each team in the amount of about $12M and required each to build and test a BDM to demonstrate “critical technologies”
- Algorithms
- Database loading/updating
Each team was provided Development Data Sets (DDS) for testing
DDS contained 120,000 subjects (here, subject = 10-print sets of electronic images scanned from a standard paper ten-print card), and about 100 mated subjects
The BDM needed a “Quick & Dirty” latent workstation that allowed manual feature encoding and searching
Each team was required to build a System Architecture simulation model
Each team given 18 months to demonstrate critical technologies
The BDM Test

- Government produced test data for use on the BDM systems
- BDM Data Set was tightly controlled by Government
  - Two day load under Government supervision
  - Supervised scrub following BDM
- The BDM Data set contained a 600k subjects to be used as the gallery
  - 500k randomly selected ten-print records (scanned paper cards) ("horizontal slice")
  - 100k subjects with similar fingerprints ("vertical slice") to measure conversion/extraction/loading into database (see Database issues below):
    - 26k arches
    - 74k small count loops (which are difficult to differentiate from arches)
- BDM testing lasted ten days under continuous Government supervision
The BDM Test (continued)

- The search (probe data) included:
  - A ten-print search set of 3,200 subjects
    - About 1800 had mates in the background database
    - Over 200 had more than one mate in the background database
      (several had up to 12 mates in the background)
  - A latent search set of 300 subjects
    - All latents had mates in the background database
    - Broadly divided into Good, Bad, and Ugly categories

- The DDS was not used in any of the Government tests so as to avoid training issues or potential gaming by the offerors
Nature of the latent matching problem

- Latent fingerprint quality extremely variable
  - No common standard
  - Latent performance numbers “hyped” by vendors, users
- Development of so called “AFIS searchable quality” (pre-IAFIS) concept – classic case of cooking the data
  - At time of print collection
  - At time of search submission
- Unavailability of sufficient test data (latent prints) makes computation of performance metrics very difficult
- Effective “miss analysis” is very difficult – ground truth problem
- Difficulty in obtaining data because of legal (evidentiary) and privacy concerns
Testing the Latent Matching Capability

- 300 latents were used to test the latent end-to-end performance
  - 3 teams of 3 latent examiners encoded each of the 300 latents on each of the 3 BDM systems (2700 encodings)
  - The latent examiners were hired and trained by the vendor
  - Latent performance varied considerably
- 30 latents and 30 rolled mates were manually encoded to obtain minutia ground truth by a team of FBI examiners
  - Analysis provided insight into algorithm performance
  - Minutia extraction accuracy
- BDM tests showed that vendors performance in encoding accuracy did not mirror their performance in matching accuracy*
  - Testing process provided a basis for evaluating minutia extraction algorithms
  - Provided basis for conclusion that matchers could be improved
  - Used for subsequent algorithm improvement

* Interestingly, MINEX Report, March 6, 2006 shows this is still true
# Subjective Quality Definitions

<table>
<thead>
<tr>
<th>Quality/Characteristic</th>
<th>Good</th>
<th>Bad</th>
<th>Ugly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ridge structure</td>
<td>Well defined</td>
<td>Some fragmentation, but most minutia relationships discernable</td>
<td>Highly fragmented, minutia relationships difficult to discern</td>
</tr>
<tr>
<td>Fingerprint image area discontinuities</td>
<td>Well defined</td>
<td>Some parts of print area is missing or blotchy, but core area is visible</td>
<td>Print image not well defined for much of print or core area poorly defined</td>
</tr>
<tr>
<td>Fingerprint likely to overlap rolled fingerprint</td>
<td>Print captures most of the core area</td>
<td>Part of the core area is present in the print</td>
<td>Core area is missing or is undefined</td>
</tr>
</tbody>
</table>
Distribution Varies by Agency

• FBI is often the processor of last resort – the latents sent to the FBI are the ones not matched by the States.
BDM Lessons Learned

- All 3 AFIS vendors dramatically improved their products because of the BDM,
- The test demonstrated good automated fingerprint classifications, for all bidders
- Excellent ten-print search performance (accuracy and response)
- Binning problems were resolved by all vendors
- Latent search performance did not meet expectations
  - Better understanding of the latent search issues
  - All vendors proposed latent performance improvement plans
Post BDM Latent analysis

- Second stage of the AFIS development was won by the Lockheed Martin Team
- Government took active part in system development
  - Design oversight
  - Algorithm testing
- Development of the Latent Ground Truth test data set
  - At least 3 latent examiners encoded each of the 300 latents and their corresponding ten-prints (gallery)
  - A comparison tool developed to find encoding differences between the latents and the gallery set for each examiner’s encoding
  - Differences resolved using group approach
- Process repeated showing only minutia visible on the latent and the file print – Ideal Latent Test Data Set
- Ground truth data was used for testing algorithm development
- The set now has 265 mated subjects and is also known as NIST Special Database 27
Development of Latent Analysis Tools and Strategies

- Tools to measure minutia extraction accuracy
- A method for estimating latent search performance for a fully populated database
- Strategies to limit gallery size (search space) using soft biometrics
- Evaluating latent quality impact on matcher performance
How and why we measured minutia extraction accuracy

- Developed and used comparison tool to find differences between automatically extracted minutia and ground truth minutia
  - % true minutia recovered
  - % false minutia produced
- Following minutia extraction algorithm modifications (and parameter adjustments) tested impact of feature extraction performance on matcher performance
- Optimized feature extraction algorithm performance as part of the end-to-end search process
- Identified problem areas and tested potential solutions
  - High curvature areas
  - Exaggerated impact of false minutia
How to estimate performance for a fully populated database?

- Used various size databases to test impact on performance as a function of gallery size
  - 33k
  - 300K
  - 3M
- Validated Rae Moore’s Laws
  - If the latent is going to be found it is likely to be in the top rank about 80% of the time
  - The system FAR (selectivity) is a linear function of the database size for all large databases
- Able to use small test database for rapid algorithm improvement
How and why we limited gallery size (search space)

- Using a simulation Model we found that searching 1,000 latents per day would require about 3 times the computing resources as processing 45,000 ten-print searches.
- Decision was made to use all available physical descriptors to limit latent searches so that no more than 25% of gallery would be searched.
- For high priority searches, full gallery would be searched with administrator permission.
- Descriptors included:
  - Finger number (or range of finger numbers)
  - Other descriptors of suspect (sex, race, approximate age, etc)
- Average search space was found to be about 11-14%
How to measure latent quality impact on matcher performance?
Lessons Learned - Latent Reliability (single finger)*

*Fully populated database and using latents with quality similar to BDM set
Additional findings for latent reliability improvements

- Use of multiple latent impressions and multiple fingers shows great promise
  - 2 finger search reliability could exceed 75%
  - Multiple impressions for same case can use search result fusion to increase performance
- Top 1000 candidate reliability is potentially over 70%
  - Suggests use of automated candidate elimination algorithms
- Greater ten-print area and more minutia improve performance
- Better feature extraction algorithm
- Better matching algorithms