Forensic Science and Biometric Systems: an Impossible Mix?

With contributions from Flore Bochet and Damien Dessimoz
Challenge 1: To have a shared understanding of the concept of operations

(e.g. face recognition in police operations with a COTS system)
Selection of usable facial case data according to the planned forensic usage

Range of situations where facial information are available
- 3D to 2D
- Decrease of selectivity

Local: against a limited set of individuals (even 1:1)

Contextual: against a set of individuals known for activities in a given context (aka metadata binning)

Global: against all putative sources (similar to DNA or fingerprints)

Dissemination: facial images distributed to the press or target groups

Different customers and expectations
Different attitude towards risk
Different metrics for performance

Courtesy: Damien Dessimoz
Le même braqueur a-t-il encore frappé?

Mi-janvier, un homme braquait l’UBS de Delémont à visage découvert. Lundi matin, c’était la Banque Coop à Bienne. Les polices bernoise et jurassienne pensent qu’il s’agit de la même personne.
Global: Gallery 60k, ranked 32
Contextual Gallery selected based on MO, ranked 5. Not in a reasonable hit list with a gallery of 60k.
Local: ID documents: Gallery: 60k, ranked 3
Towards an evaluative report?

- To get a likelihood ratio for the forensic findings (E, here the score), we need two probability densities.

The within-source variability is unknown in most cases.
Towards an evaluative report?

- Does it all come down only to a judgment based on the training and experience of the expert applying the recommended methods?

- Without data on the within-source variability, how can an expert robustly assign a likelihood ratio?
Challenge 2: To have a clear definition of the respective role of technology and forensic experts

(e.g. lights-out ID)
Lights-out Decisions

- Results of Flore Bochet on a set of 1818 marks
- Auto-encoded without any other user input
- Searched against a COTS AFIS system (background database of about one million fingerprints)
Auto-encoding and Search

Rank [1]: 69%
Ranks [1:10]: 71%
Ranks [1-20]: 72%
Ranks [1-50]: 72%
Ranks [1-100]: 73%
Ranks [1-500]: 74%
Auto-encoding and Search

• In about 70% of the cases, the identification will be made after checking (verifying) only rank [1] without any manual encoding.

• Allow to concentrate the efforts on the remaining 30%.

• Gain obtained allows to increase the number of submissions.
Providing intelligence?

- Could we use an **AFIS in lights-out mode** to provide intelligence information in the form of investigative leads delivered in a timely manner?


- **Each transaction comes with a set of variables directly from the COTS AFIS system**
  - Score at rank 1
  - Score for the candidate at rank 2
  - The number of encoded minutiae
  - Measures of quality (global and for minutiae)
  - The general patterns (mark and print)
  - The finger number of the print

![Bar chart](chart.png)

- **Correct Source**
- **Wrong Source**

<table>
<thead>
<tr>
<th>Rank [1]</th>
<th>Number of marks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1259</td>
</tr>
<tr>
<td></td>
<td>559</td>
</tr>
</tbody>
</table>

True state regarding the candidate returned at Rank 1

Rank [1]: 69%
**Random Forest classification**


<table>
<thead>
<tr>
<th>Difference between score @Rank 1 and @Rank 2</th>
<th>Importance</th>
<th>Error when added to the model</th>
<th>Drop of the error with the addition of each variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score @Rank 1</td>
<td>168</td>
<td>0.11</td>
<td>0.06</td>
</tr>
<tr>
<td>LFIQ (Yoon &amp; Jain 2012)</td>
<td>60</td>
<td>0.10</td>
<td>0.01</td>
</tr>
<tr>
<td>All other variables</td>
<td>&lt;33</td>
<td>&lt;0.10</td>
<td>≈ 0</td>
</tr>
</tbody>
</table>
Balancing risks

How many misleading leads the police force is ready to cope with?

Assume about 340 marks submitted per week. 2% misleading information could be viewed as fit for purpose for your police force.

<table>
<thead>
<tr>
<th>True state</th>
<th>Predicted state</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct source</td>
<td>134</td>
<td>90</td>
</tr>
<tr>
<td>Wrong source</td>
<td>3</td>
<td>116</td>
</tr>
</tbody>
</table>

About 130 instant IDs per week for free

For 3 misleading info per week

How many misleading leads the police force is ready to cope with?
Challenge 3: To understand how biometric systems can help with forming evaluative opinions

(e.g. LR systems made available to the fingerprint community)
Recent changes in reporting

“Individualization of an impression to one source is the decision that the likelihood the impression was made by another (different) source is so remote that it is considered as a practical impossibility”

2.45 - The *decision* whether or not a mark can be individualised is potentially a complex one calling for a series of subjective judgments on the part of the examiner. The *decision* is one of opinion, not fact.
38.24 - What matters **more than the choice of language** (whether the witness says that he is ‘confident’, ‘sure’, ‘certain’ or ‘in no doubt’) is the **transparency** of the opinion.
Transparent decision process

Weight of the forensic findings (MP < 10^{-15})

ID decision – An adventitious match is a practical impossibility

“Leap of faith” (D. Stoney)

1. Who left the mark? (priors)
2. Fingerprint Evidence (LR)
3. Who left the mark? (posteriors)

1/all fingers
LR for FP comparison 99.99...%

Utility

Equilibrium in the *spectrum* of knowledge

- Relevant systematic studies (published or documented)
- Structured portfolio of cases
- Proficiency and collaborative testing
- Years of experience
- Unstructured data collection from uncontrolled casework

**Probabilistic** vs **unfettered opinions**
Arrival of probabilistic models

• Help assign the weight of evidence to the whole configuration without decomposing the contribution of its individual minutiae.

• The more recent efforts have been successfully presented to the Royal Statistical Society: C. Neumann, I. W. Evett, and J. Skerrett, "Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm," *Journal of the Royal Statistical Society*, vol. 175, pp. 371-415 (with discussion), 2012.
Neumann & al. (2012)

(a)

(b)

Fig. 2. Illustrations of the radial triangulation used to order minutiae features that are recorded in a feature array. A feature array that is recorded from a randomly selected fingerprint provides the notation for the data that are extracted from the minutiae video. Furthermore, the circular acquisition of the minutiae with respect to the centre, the capture of information for a given configuration is invariant to distortion. For a given configuration the minutiae (the variables listed in Table 1 are indicated in grey in (a)); the differences between (a) and (b) are presented by using thinner lines.
Forensic Error Rates

What is an acceptable error rate?

We should confirm them through an operational validation

Models can be used..

41.32 [...] to provide background data to assist fingerprint examiners with their evaluation of marks and to enable them to express the strength of their conclusion in a transparent and verifiable manner.

The situation we want to handle

My opinion is that the mark has been identified to the right thumb of Mr X.

The probability for the mark to originate from someone else is so small that I consider it to be a practical impossibility.

We submitted your case a statistical analysis through the University XYZ, Prof S. Tat

The LR obtained is 1.8e+6, that amounts to a match probability of 5.6e-7

How do you get from a *match probability* of 5.7e-7 to an *identification*?
We need to precisely define how these “experts” will operate.

A set of SOPs need to be drafted before any operational implementation.

To embrace it, I need to understand and trust the model.

I need to be able to explain its meaning and limitations.

Models are inevitable in the future.

As in DNA, probabilities will be asked by both prosecution and defence.

I may want to identify regardless of the number given by the model.

Examiners’ feedback.
Conflict resolution procedures

1. First examiner
2. Statistical Model
3. Decision
4. Second examiner (verifier)
Understanding Our Decisions

- Understanding the concept of “sufficiency” in friction ridge examination

  C. Neumann, C. Champod, M. Yoo, G. Langenburg, T. Genessay, NIJ award - 2010-DN-BX-K267

- Results presented at the annual meeting of the American Academy of Forensic Science, Feb 2013

- 15 comparisons (chosen to highlight decision boundaries)
  - 12 pairs latent/control prints from same source
  - 3 pairs latent/control prints from different sources

- All annotations captured through a web-based software designed to support the ACE process (Picture Annotation Software – PiAnoS – https://ips-labs.unil.ch/pianos/index.html)

- Approximately 600 examiners contacted
  - 145 completed first comparison
  - 123 completed all 15 comparisons
Data from C. Neumann, C. Champod, M. Yoo, G. Langenburg, T. Genesssay, Understanding the concept of “sufficiency” in friction ridge examination NIJ award - 2010-DN-BX-K267
User 342 (ID) – not certified (3 years)

LR = 3.17e-18
User 436 (ID) – Certified (5 years)

LR = 9.03e-22
User 481 (ID) – Certified (7 years)

LR = 723657
Conflict resolution procedures

First examiner

Statistical Model

Second examiner (verifier)

Decision
<table>
<thead>
<tr>
<th>Holistic expert</th>
<th>LR-based biometric system</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level 3 features</strong>&lt;br&gt;(pores and ridge edges)&lt;br&gt;(scars, creases)</td>
<td><strong>Level 1 features</strong>&lt;br&gt;(general flow/pattern)</td>
</tr>
</tbody>
</table>

Their contributions must be articulated through an **argumentative discourse** used to assign the numerator and denominator of the **likelihood ratio** associated with the features **not covered** by the biometric system. To the very least an **error rate** obtained from task-relevant proficiency tests should be disclosed.

**Likelihood ratios** assigned following a documented and systematic account of the **within** source and **between** sources variations.
Saying more than “inconclusive”

Again, we need to precisely define the scope of usage.

Examiners’ feedback:

- The statistics may convey more weight than it deserves.
- Very useful source of additional information, either as evidence or for intelligence purposes.
- We don’t want to mislead anyone.
- Already a reality for the Dutch NFI.
Diminishing match probabilities (or increasing LR)

- [1 – 1/10^3]
  - Risks of misleading may outweigh benefits
  - Need more resources

- [1/10^3-1/10^-9]
  - Helps with the decision making
  - Intelligence tool
  - Need more resources

- [<10^-9]
  - Allows transparency but marginal impact on decision making
  - Actual resources
Biometric (Mars) and forensic (Venus) wedding?

① Friendly visit to the other clubs/planets
② Suggested elements of the *prenuptial agreement*:

- Jointly define the *use cases* that address the needs of forensic investigation.
- Jointly agree on the *territories of excellence* (lights-out versus manual operations).
- Jointly identify the mutual benefits in relation to the production (the birth) of *evaluative reports*. 
Contact details

Prof. Christophe Champod
Ecole des sciences criminelles / Institut de police scientifique
Batochime / Quartier Sorge
CH-1015 Lausanne

Tel: +41 (0)21 692 46 29
Fax: +41 (0)21 692 46 05
E-mail: christophe.champod@unil.ch