Signature Biometrics

Prof. Julian FIERREZ

Universidad Autonoma de Madrid - SPAIN

http://atvs.ii.uam.es/fierrez

Funding Acknowledgements

Public

Private

Introduction

- Signature is one of the most socially accepted biometric traits, it has been used for centuries to validate legal and commercial documents and transactions.
- Automatic signature recognition has some general challenges:
  - **Large intra-user variability** (behavioral biometric, inter-session)
    - Difficult to model, large amount of training data (usually scarce)
  - Small inter-user variability (in case of forgeries)
    - The skill level of actual forgeries is unpredictable

![Signatures from the same user](image)

- High variability
- Low variability

![Skilled Forgery](image)
Introduction

- **High deployment** of multiple electronic devices
- Signatures can be easily captured by means of multiple devices
- **High deployment** in banking and commercial sectors

---

Biometric Market by Modality

- Decreasing (in *Relative Importance*): **Fingerprint**, from 48% to 15% (31% w AFIS)
- Growing: **Iris** from 9% to 16% and **Face** from 12% to 15%
- Huge grow: **Speech** from 6% to 13% and **Signature**, from 2% to 10%
Behavioral Biometrics

- Human activity patterns are clearly established from childhood
- As patterns, they are stable and reproducible, though subject to variability
- Neuromotor coordination of gestures and movements
- Continuous identity monitoring possible
- User is an active part of the play
- Multilevel strategy: from dynamic trajectories to expressions, context, habits, stylometry, experiences
- Not fixed patterns but changing and adapting ones

Active Authentication by DARPA
Signature as Behavioral Pattern

- Human interaction permits transparent authentication
- Make use of existing input channels, no added specific sensors:
  - Handwritting (tablets and pads)
  - Mouse dynamics
- Other sources of variability (sensor, session) included into behavior pattern modelling / compensation
- Fully revocable patterns
- Incorporates soft biometrics (gender, handedness, language, ...)
- Easy of use, high user acceptance

Signature Recognition


On-line Signature Verification: Overview

Feature-based (Global Features)
- Distance-based classifiers
  - Mahalanobis
  - Euclidean \([\text{Nelson et al., 1994}]\)
- Statistical/other classifiers
  - Gaussian Mixture Models (GMM)
  - Parzen Windows

Function-based (Local Features)
- Time-Sequence matching techniques
  - Hidden Markov Models (HMM) \([\text{Dolfing et al., 1998}]\)
  - Gaussian Mixture Models (GMM) \([\text{Richiardi et al., 2005}]\)
  - Dynamic Time Warping (DTW) \([\text{Sato and Kogure, 1982}]\)


On-line Signature Verification: System Model

1. Data Acquisition & Pre-Processing
2. Feature Extraction
3. Similarity Computation (Matching)

**Signature Acquisition: Input Data**

- **Time resolution:** 100-200 samples/sec
- **Space resolution:** 1000 pixels/inch resolution

Measured:
- x, y coordinates of the signature trajectory
  - on pen down
- time stamp at each sample point
- pressure at each point
- pen inclination angles at each point
  - altitude (0-90)
  - azimuth (0-359)
- ...

**Signature Pre-Processing**

Reduce sensor interoperability issues due to diverse devices and writing tools (stylus/finger)

- Size normalization and centering
- Pressure normalization
- Resampling

**Pre-Processing: Re-Sampling**


**Feature Extraction**

Feature Extraction: Global Features


Feature Extraction: Global Features Example
Feature Extraction: Global Features Example


Global Features: Performance (on MCYT DB)

<table>
<thead>
<tr>
<th></th>
<th>5 training signatures</th>
<th>20 training signatures</th>
</tr>
</thead>
<tbody>
<tr>
<td>SKILLED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RANDOM</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Feature Extraction: Time Sequences

<table>
<thead>
<tr>
<th>#</th>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>x-coordinate</td>
<td>(x_t)</td>
</tr>
<tr>
<td>2</td>
<td>y-coordinate</td>
<td>(y_t)</td>
</tr>
<tr>
<td>3</td>
<td>Pre-pressure</td>
<td>(p_t)</td>
</tr>
<tr>
<td>4</td>
<td>Path-tangent angle</td>
<td>(\delta_t = \text{arctan}(p_t/x_t))</td>
</tr>
<tr>
<td>5</td>
<td>Path velocity magnitude</td>
<td>(v_t = \sqrt{p_t^2 + x_t^2})</td>
</tr>
<tr>
<td>6</td>
<td>Log curvature radius</td>
<td>(\rho_t = \log(1 \times x_t) = \log(g_{x_t}, \delta_t)), where (x_t) is the curvature of the pen's trajectory</td>
</tr>
<tr>
<td>7</td>
<td>Total acceleration magnitude</td>
<td>(s_t = \sqrt{x_t^2 + y_t^2} = \sqrt{\Delta x^2 + \Delta y^2}), where (s_t) and (v_t) are respectively the tangential and centripetal acceleration components of the pen motion.</td>
</tr>
<tr>
<td>8-14</td>
<td>First-order derivative of features 1-5</td>
<td>(x_{t,n}, y_{t,n}, \rho_{t,n}, \theta_{t,n}, \gamma_{t,n})</td>
</tr>
<tr>
<td>15</td>
<td>Pen azimuth</td>
<td>(\gamma_t)</td>
</tr>
<tr>
<td>16</td>
<td>Pen altitude</td>
<td>(\delta_t)</td>
</tr>
<tr>
<td>17-18</td>
<td>First-order derivative of features 15-16</td>
<td>(x_{t,n}, y_{t,n}, \rho_{t,n}, \theta_{t,n}, \gamma_{t,n})</td>
</tr>
</tbody>
</table>


Feature Extraction: Time Sequences

**Similarity Computation**


**Dynamic Time Warping**

**Dynamic Time Warping**

\[ D(i,j) = \min \begin{cases} \ D(i-1,j-1) + d_e(i,j) \\ D(i-1,j) + d_e(i,j) \times c \\ D(i,j-1) + d_e(i,j) \times c \\ d_e(i,j) < \text{thresh} \rightarrow 0 \end{cases} \]

\( D \) serves to define the optimal alignment between point \( i \) in the input signature and point \( j \) in the template, which is computed via **dynamic programming**.

A constant factor \( c \) multiplied by the Euclidean distance between the two feature vectors is used instead of constant penalties.

No penalty if the Euclidean distance is small.


---

**Stochastic Approach: Gaussian Mixture Models**

- Probability of occurrence modeled through a mixture of Gaussians
- Model constructed with several training samples to incorporate sample variability
- Compact representation

Performance Evaluation: Signature Databases

- Databases allow **systematic evaluation** of algorithms
- **Large** publicly available databases are **scarce**, mainly due to
  - **Legal and privacy** issues
  - **Huge resources** needed to capture and process the data
- **MCYT database** has been the most widely used dataset since 2003, reaching performances on 330 subjects below 1% ERR
- **Other existing databases** include SVC, Biomet, MyIdea, Susig
- Recently, new databases containing **additional features** have been captured (e.g., BioSecure Multimodal Database, e-BioSign)


Traditional Acquisition Scenario (2000-2015)

**Benchmarks: SVC 2004**

- **Challenging data:**
  - WACOM Intuos pen tablet with inkless pen (i.e., without visual feedback).
  - Invented signatures different to the ones used in daily life.
  - English and Chinese signatures.
  - Impostors know the dynamics of the signatures being forged.

- **Acquisition protocol:**
  - 40 subjects.
  - 20 genuine signatures (2 sessions) + 20 skilled forgeries (from five impostors)

### Benchmarks: SVC 2004

**SVC-04 skilled forgeries**
- Average ROC for 10 genuine signatures and 20 skilled forgeries

**SVC-04 random impostors**
- Average ROC for 10 genuine signatures and 20 random impostors

Our HMM system
- DTW from Turkey


---

### Resources: Multimodal Databases w Signature

- **MCYT Database** (Spanish Project 2000-2003)
  - Fingerprint (with human-labeled quality) and on-line Signature of 330 donors

- **BiosecurID Database** (Spanish Project 2003-2006)
  - 8 Modalities: speech, iris, face, Signature and handwriting (on-line and off-line), fingerprints, hand and keystroking of 400 donors in 4 acquisition sessions

- **Biosecure Database** (EU Project 2004-2007)
  - 3 Datasets: Web scenario, Office scenario, Mobile scenario
  - 667 donors

See: https://atvs.ii.uam.es/atvs/databases.jsp


Acquisition Example: MCYT Signature

- Acquisition procedure:
  - WACOM Intuos pen tablet.
  - Ink pen over paper → both online and off-line corpus.
  - Restricted size grid guidelines.

- Acquisition protocol:
  - 330 subjects.
  - 25 genuine signatures (five sessions) + 25 skilled forgeries (from five impostors)

Acquisition Example: Biosecure Multimodal DB

- PHILIPS SPC 900NC + PLANTRONICS Voyager S10
- LG IrisAccess EO3000
- BIOMETRIKA FX2000
- YUBEE (Atmel FingerChip)
- WACOM Intuos A6 + Inking Pen
- CANON EOS 30D + Ring Flash
Examples from Biosecure MDB

Tablet

Mobile

Benchmarks: BSEC 2009

- DTW, HMM and Global Systems
- Score normalization
- Fusion of systems

### Benchmarks: BSEC 2009 - Forgeries

- *Shoulder surfing* (visual access to drawing process)

![Graphs showing False Acceptance Rate vs False Rejection Rate for different sessions.](image)


### Template Aging in Signature

- 29 common users from BiosecureID and Biosecure.
- 6 sessions with a 15-month time span (inter-session).
- 46 genuine signatures: 4 + 4 + 4 + 4 + 15 + 15
- 10 skilled forgeries per user

![Timeline showing sessions S1 to S6 with time spans and aging process.](image)

Examples of the multi-session DB


Examples of the multi-session DB

Fixed template, varying test

- Mean genuine score evolution: significant template drift (>6 months)
Template Update: Fixed test, varying enrollment

Reference: 12 months (4 sign.)
Complete update (4 sign.)
Mixed update (4 + 4 sign.)
Complete update (8 sign.)

Compared to the reference scenario (12 months train-test):
• Significant improvement by forgetting and retraining using a small set of new training data.
• This can be further improved by not forgetting but adapting using the new data.
• Enough new train data available  better than using old data.

DATA-DEPENDENT PROBLEM, STRONGLY DEPENDENT ON THE AMOUNT OF TRAINING DATA

More on Biometric Aging and Template Update

Performance in 2015 → BIOTRACE100

• **Accuracy (Signature Long-Term - SLT Database):**

<table>
<thead>
<tr>
<th></th>
<th>4 training signatures</th>
<th>16 signatures</th>
<th>31 signatures</th>
<th>41 signatures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Forg.</td>
<td>97.2 %</td>
<td>99.3 %</td>
<td>99.9 %</td>
<td>99.9 %</td>
</tr>
<tr>
<td>Skilled Forg.</td>
<td>88.3 %</td>
<td>93.1 %</td>
<td>95.9 %</td>
<td>99.3 %</td>
</tr>
</tbody>
</table>

• **State of the art performance**

• **Template and system configuration update strategies in order to minimize the aging effect**


Banking Industry - Tech Transfer to

• Stylus and finger-drawn signature recognition
• Off-line fraud detection and on-line verification
• Semi-automatic tools to aid experts in signature comparison (lawsuits)

Dynamic signature acquisition and management solution already in operation (> 46k sensors, > 500M operations/year)

- 70 users, 2 capturing sessions. 5 devices (4 Wacom, 4 Samsung)
- 8 genuine signatures and 6 skilled forgeries per user and device
- Stylus and finger as writing tools (Samsung)


2017 Performance on e-BioSign (Modern Devices)


From Signature to Touch Gestures

- Graphical Password-based User Authentication with Free-form Doodles


Graphical Passwords

- Gesture-based authentication on touch-screens
- Slow typing in touchscreens
- Biometric-rich gestures
- Revocability

Graphical Passwords: Related Works

1. Draw a Secret [Jennyn et al., 1999]
   US Patent 8024775 B2

2. Pass-Go [Tao et al., 2008]

3. Pattern Lock [Google]
   US Patent 20130047252 A1

4. Picture Gesture Authentication [Microsoft]
   US Patent 20130047252 A1

5. Multi-touch gestures [Sae-Bae et al., 2012]
   US Patent 2013019490 A1

Graphical Examples

- Doodles
  - Genuine samples
  - Forgeries

- Pseudo-signatures
  - Genuine samples
  - Forgeries

Some Initial Results

- Verification performance on the validation set
- Just x, y features
- Score fusion of GMM and DTW


Current Work: Swipe Biometrics

- Continuous user authentication through touch biometrics:
  - Security beyond the entry-point
- Situation:
  - Freely interacting with the touchscreen while reading or viewing images
Current Work: Swipe Biometrics

Other Recent Advances:
Synthetic Signature Generation

- Novel signature generation schemes using **data-driven spectral features**, or **human neuromotor properties**, which generate realistic yet random full X, Y, and Pressure signature signals.
- Useful for improving the training with limited data.


Other Recent Advances: Template Protection

- Biometric data can be compromised if raw signals are stored

- **Template protection schemes** needed to secure user privacy
  - **Biometric cryptosystems**: combination of cryptographic keys and biometric data (e.g., fuzzy vault, fuzzy commitment)
  - **Transform-based schemes**: application of non-invertible functions to the biometric data (e.g., cancelable biometrics)

- **Dealing with variability** is the main challenge in this field


---

The Future of Behavioral Biometrics

**Challenge 1: Adapting to New Application Scenarios**


The Future of Behavioral Biometrics
Challenge 2: Incorporating Contextual Information


The Future of Behavioral Biometrics
Challenge 3: Adapting to the User (e.g., Aging)

The Future of Behavioral Biometrics
Challenge 4: Exploiting Big Data


Signature Biometrics: Conclusions

- Revocability
- User intra-variability
- Easy of use, user acceptance
- Multi-sample training
- Less sensor-interoperability issues
- Model updating
- Easy to integrate at low-cost
- Multilevel strategies
- Continuous ID
- Data scarcity

- Mature technology
- Major role in on-line, mobile, and legacy applications
- User convenience to drive application development
- Room for substantial industry-applicable research
Signature Biometrics

Prof. Julian FIERREZ

Universidad Autonoma de Madrid - SPAIN

http://atvs.ii.uam.es/fierrez