Toward Intelligent Enrollment in Biometrics, with Case Studies in Keytroke and Signature Biometrics

Dr. Julian Fierrez

Biometric Recognition Group – ATVS
Escuela Politécnica Superior
Universidad Autónoma de Madrid, Spain

(Visiting scholar at)
Pattern Recognition and Image Proc. Lab.
Dept. of Computer Science & Engineering
Michigan State University

Index

• The presentation is structured as follows:

  Introduction: Preliminaries and problem statement

  Related work: The biometric zoo, exploiting the biometric zoo

  Case Study: User verification based on keystroke dynamics

  Case Study: PDA-based online signature verification

  Conclusions

Recent advances in online signature verification
Introduction: Preliminaries

- Biometric system: Automatic pattern recognition system that makes use of personal biometric traits to recognize individuals
  - Enrollment
  - Verification (Authentication): 1-to-1 matching
  - Identification: 1-to-N matchings
Introduction: Preliminaries

- Biometric verification is a detection task:
  - Type I Error, **False Rejection** (FR): a genuine user is rejected
  - Type II Error, **False Acceptance** (FA): an impostor is accepted
  - **Equal Error Rate** (EER): error rate for the decision threshold where FA = FR

- A more general architecture of multimodal biometric verification with score-level fusion including quality measures (**quality-based biometrics**):
  - Q-based enhancement
    [Hong et al. 98]
  - Q-based feature weighting
    [Chen et al. 05]
  - Q-based fusion
    [Bigun et al. 97, 03]
    [Fierrez et al. 05, 06]
    [Nandakumar et al. 06, 08]
  - Failure to acquire event
    [Simon-Zorita et al. 03]
    [Chen et al. 05]
Introduction: Problem statement

- A related (but not equivalent) problem:

    ![Diagram of enrollment and verification process]

    **Intelligent Enrollment:**
    Exploitation of the enrollment data (multiple samples) not only to create the templates/models but also to adjust in a user-dependent way some parameters of the system during verification

Introduction: Intelligent enrollment (example)

- Average NFIQ [Tabassi et al., ICIP 2005] of the enrollment fingerprints (typically just one) to reject problematic users: single fingerprint → single output $Q$ as a performance prediction

    ![NIST Fingerprint Image Quality](image)

    - Works well with fingerprints, but not adequate for biometrics in which the discriminative power is strongly dependent on the user intra-variability (most behavioral biometrics)

- Our (more general) proposal: multiple enrollment samples of a given subject → one/multiple outputs (e.g., user performance prediction ~ user $Q$) → user-dependent system adaptation → Why bother with user-dependent adaptation?
Related Work

- The biometric zoo
- Exploiting the zoo

Related Work: The biometric zoo

- The users of a biometric system can be classified into different groups depending on their individual performance → the biometric zoo

- This classification has been well studied in behavioral biometrics such as speaker (e.g., NIST SRE campaigns [Doddington et al., ICSLP 1998]). A few studies exist also in more widespread biometrics such as face (e.g., [Wittman et al., CVPR 2006]) and fingerprint (e.g., Hicklin et al., NISTIR 2005)

- This classification is nowadays even more relevant because of new behavioral biometrics (keystroke dynamics, mouse dynamics, ...), and new challenging acquisition devices to which some users may not be habituated (e.g., smart phones), causing devastating effects if not properly handled
Related Work: The biometric zoo (speaker)


- **Sheeps**: good for recognition systems (most of the population)
- **Goats**: difficult to recognize by an automatic system ($\rightarrow$ FR)
- **Lambs**: very similar one from another and/or easy to imitate ($\rightarrow$ FA)
- **Wolves**: great imitators ($\rightarrow$ FA)

Additions to the zoo (not speaker):

[Bolle et al., PR 2002]
- **Chameleons**: easy to imitate and good imitators

[Richiardi et al., Biosecure 2008]
- **Antelopes**: able to avoid the attack of impersonators

---

Related Work: The biometric zoo (signature)

[Yeung et al., “SVC 2004”, ICBA 2004], [Fierrez et al., TSMC-C 2005]:

---
Related Work: The biometric zoo (fingerprint)

- The ubiquitous 2% [NIST, "Summary of NIST standards for biometric accuracy, tamper resistance, and interoperability", NISTIR, 2002].

- [Hicklin et al., "The myth of goats: how many people have fingerprints that are hard to match", NISTIR 7271, 2005].

"The definition of a Goat, or person whose fingerprints are intrinsically hard to match, varies. However, results clearly show that the proportion of Goats is very small, regardless of the definition. None of the 6,000 subjects had fingerprints that were always hard to match (with single-finger match scores worse than a threshold corresponding to a verification False Accept Rate of 1%), less than 0.05% of the subjects had fingers that were usually hard to match; less than 0.3% of the subjects had fingers that were hard to match even a quarter of the time. Many individuals were particularly easy to match: for 77% to 81% of subjects, every fingerprint comparison had match scores better than a threshold corresponding to a verification False Accept Rate of 10⁻⁶ (0.0001%)."

Related Work: The biometric zoo (face)

- Analyzing the zoo
- Exploiting the zoo
Related Work: The biometric zoo (face)


![Goat Plot](image1)

Figure 1: Goat Plot

- From [Kang et al., "Continual Retraining of Keystroke Dynamics Based Authenticator", ICB 2007].

Related Work: The biometric zoo (keystroke)

- User clustering and classification, user quality, enrollment policies and exception handling, ...

![Goat Plot](image2)

Figure 2: Goats

Distance between enrollment and login patterns

Sheep - Sheep - Goat - Unstable Sheep-Goat

Template Update/Adaptation

![Goat Plot](image3)
Related Work: The biometric zoo, observations

- Menagerie dependent on: biometric, database, acquisition scenario, population, training/testing conditions, ... \(\rightarrow\) DATA-DRIVEN problem!!
- Generalization (does the ubiquitous 2% makes sense??) only reasonable in large-scale experiments, usually government-sponsored evaluations (>1,000s of subjects). Anyway, the generalization is linked to particular technologies
- Zoo assessment is an open and important problem for large biometric deployments (face, fingerprint, ...), only accessible for academic research once large DBs are available (e.g., NIST FRGC, NIST fingerprint-face multimodal biometric scores*, etc.)
- Once the variables are fixed (the data-driven problem is stated), clustering techniques may be useful to assess the biometric menagerie
  \(\rightarrow\) Can we exploit the zoo? (much more accessible for academic research)

* "... The data is intended to allow ... the study of score-level fusion ... and assessment of the existence of the biometric zoo."

Related Work: Exploiting the zoo (UD processing)

- UD features
  [Fairhurst et al., IPRAI 94]
- UD modeling
  [Martinez et al., ICFHR 08]
- UD score normalization
  [Fierrez et al., TSMC 05]
  [Poh et al., MMUA 06, TASLP 08]
- UD fusion
  [Jain et al., ICIP 02]
  [Toh et al., TSP 04]
  [Snelick et al., PAMI 05]
  [Fierrez et al., PR 05]
- UD decision
  (e.g., UD thresholds [Jain et al., PR 02], failure to enroll events and exception handling)
Related Work: Exploiting the zoo

[Hocquet et al., “User Classification for Keystroke Dynamics Authentication”, ICB 2007]:
- 31 global features (duration, statistics of time vectors Press-Release, PP, RP, RR) → PCA → 5 features → k-means clustering → 3 classes
- Recognition based on weighted sum of 3 scores (distance, disorder, time discretization) between enrolled and test time vectors
  → Cluster-dependent (UD) fusion and decision

Limitations:
- No partition development/test
- Only 38 subjects with few problematic: “... our results minimize the influence of low performance users, who have catastrophic results. We have identified three of this type of users in our base (EER>30%), but they have given only a few sequences so their influence is small.”

Case Study: User verification using keystroke dynamics

- Related work
- Proposed system
- Database and protocol
- Experimental results
Case Study (keystroke): Related work

[Peacock et al., IEEE Sec. & Privacy 2004]
- Survey of keystroke biometrics
- Largest published studies about 100 users (about 1,000 trials)
- Very heterogeneous training/testing conditions (# keystrokes)
  - Enrollment: 40 to 5,000
  - Testing: 7 to 800
- Very heterogeneous results (Total Error Rate between 1% and 50%)

[Joyce and Gupta, Comm. ACM 1990]
- Time vectors (keystroke latencies) + $l_1$ norm + (mean, std)

[Monrose et al., IJIS 2001]
- Password hardening (entropy due to keystrokes: guessing entropy)

[Bergadano et al., ACM Trans. Inf. Syst. Sect. 2002]
- Di- and tri-graphs

- Only 41 subjects but many sequences username-password (about 10,000)
- Comprehensive global feature vector
- Various modern machine learning approaches (e.g., random forests)

[Kang et al., Proc. ICB 2007]
- Continual retraining (template update)

[Hocquet et al., Proc. ICB 2007]
- User clustering → cluster-dependent fusion and decision
Case Study (keystroke): Proposed system

Enrollment

Classifier Building

Login

Password

Function-based:
- Vectors of time events (independently Press-Release)
- Exact keystroke correspondence is assumed
- Others function-based
- Based on di-graphs
- Feature-based

Distance-based:
- Modified $l_2$:
  $$\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sigma_i} |r_i - \mu_i|$$
  (better than $l_1$, $l_2$, modified $l_2$)
  (implementation trick for sigmas)
- Others depending on new feature descriptions

Identity claim

• Large DB (118 → 400) → assessment of Doddington’s zoo (→ problematic users)
• A priori problematic user detection (using only enrollment data)
• User classification and class-dependent features/matching/fusion (development/test subsets)

Julian Fierrez, PRIP Seminar, Dept. CSE, MSU – August 13, 2008
Case Study (keystroke): Database (Biosecur-ID)

- 400 individuals acquired in 7 Spanish institutions
- 8 Modalities: speech, iris, face, signature and handwriting (on-line and off-line), fingerprints, hand and keystroking
- 4 Sessions. Three levels of temporal variability:
  - Within the same session.
  - Within weeks (consecutive sessions about 2 weeks apart).
  - Within months (non-consecutive sessions)

Keystroke data:
- 4 sessions (2 to 4 weeks between them)
- 4 repetitions of the user full name + the keystroke sequence of 3 other users
  → 16 genuine sequences and 12 impostor sequences per user (worst case scenario, both users and impostors are typing exactly the same sequence)
- Real full name in Spanish (~habituated users) = single or double first name + 1 to 3 family names → 1 to 4 spaces, average number of keystrokes = 25
Case Study (keystroke): Experiments

Training with all the data from 1, 2, or 3 sessions; testing with the remaining sessions

Observations:
- The performance increases significantly from 4 to 8 training sequences (1 to 2 sessions)
- Most errors come from a set of problematic users (~20%, unhabituated to keystroking?)
- Larger training sets benefit most users with few errors but not problematic users (goats?)

→ Importance of goats in keystroking, can we detect them?

A posteriori!!!

A priori? Without any assumption about impostors (or just background information unrelated to the test data)?
Case Study (keystroke): Experiments

A priori genuine distribution $G$ using LOO during enrollment $\Rightarrow$ goatness* metrics ($\sim$ user $Q$):

- $\max(G) - \min(G)$
- $\sigma(G)$
- $\sum D_{\text{near}}$
- $\sum D_{\text{far}}$

$\Rightarrow$ goatness goodness $= \frac{1}{N} \sum | EER_{\text{apos}}(\text{apos_rank}) - EER_{\text{apos}}(\text{apri_rank}) |$

In bars the $a$ priori highest ranked users according to the $\sigma$ goatness metric (the higher the bar the higher the goatness metric), situated in their corresponding $a$ posteriori EER

Observations:
- The goatness metric improves with the training set size
- Goatness metric only useful for large training set sizes and just to detect a few of the problematic users

$\Rightarrow$ Next step: Problematic user detection as a supervised classification problem (specific features, specific classifier, binary/multi-class classification*, development/test)

* [Snelick et al., PAMI 05] Lambness metric tested $a$ posteriori

* [Tabassi et al., ICIP 05] Fingerprint Quality as a Performance Predictor
Case Study (keystroke): Experiments

- Comparison with other approaches: Database size (~ confidence in the results)

![Comparison with other approaches: Database size (~ confidence in the results)](image)

- Comparison with other approaches: Training/testing conditions

![Comparison with other approaches: Training/testing conditions](image)
Case Study (keystroke): Experiments

- Comparison with other approaches: Performance results

**Case Study: PDA-based online signature verification**

- Introduction
- Database
- System description
- Experiments
Case Study (signature): Introduction

- On-line signature verification:
  - Accept/reject identity claim
  - Dynamic information: $x$, $y$, pressure, ... vs time
- Behavioral biometric (e.g., speaker, keystroking, ...)

![Signature verification diagram]

Case Study (signature): Database

**BIOSECURE MULTIMODAL DATABASE:**

- 11 European institutions
- Three datasets (approx. 400 common subjects), 2 sessions:
  - DS1 – Internet (>900 individuals): Voice, face
  - DS2 – Desktop (>600 individuals): Voice, face, signature, fingerprint, iris, hand
  - DS3 – Mobile indoor/outdoor (>600 individuals): Voice, face, signature, fingerprint

DS2: Controlled quality data
(x, y, pressure, angles)

DS3: Degraded condition
(x, y)
Case Study (signature): Preprocessing

Signature capture process

Genuine Signature and associated signals before preprocessing

Genuine Signature and associated signals after preprocessing

Skilled Forgery and associated signals

Case Study (signature): System description

- Position and rotation normalization after preprocessing
- Feature extraction: 6 time functions (x, y, path-tangent angle, path velocity magnitude, log curvature radius, total acceleration magnitude)
- User-independent Hidden Markov Models (4 states, 16 Gaussian mixtures per state)
Case Study (signature): Protocol

- **Training phase:**
  - User models: trained using 5 genuine signatures from the first session

- **Test phase:**
  - **Genuine** scores are obtained using the remaining 15 genuine signatures (second session, about 1 month apart)
  - **Skilled forgery** scores are obtained using the 15 forgeries available per user:
    - The imitator had access not only to the visual image but also to the dynamics of the signature being imitated (player)

Observations:
- Performance in PDA is much worse than in Tablet (24% EER vs 19% EER approx.)
- The increased error is both due to a worsening of some existing goats and new PDA goats (about 25% more, for a total of about 85% of the users having some errors) → very challenging scenario (users not habituated?, too good forgeries?, time lapse?, …)
Conclusions

- **Intelligent enrollment**: exploitation of the enrollment data (multiple samples) not only to create the user templates/models but also to adjust in a user-dependent way some parameters of the system during verification (e.g., UD fusion, UD decision, user rejection and exception handling, etc.)

- Standard **sample Q measures inadequate for behavioral biometrics**

- **Assessment of the biometric zoo** for widespread modalities and application scenarios → **Exploitation of the biometric zoo** (if possible with general methods applicable to various biometrics scenarios)

- *(Take home message)* Be careful with the biometric zoo:
  - A few users may be responsible for most of your errors (e.g., keystroke dynamics) → problematic user detection and handling
  - Errors may be distributed among most users (e.g., PDA-based online signature) → improve your recognition strategy or improve your input data (e.g., more control in the acquisition)
Toward Intelligent Enrollment in Biometrics

Dr. Julian Fierrez

Biometric Recognition Group – ATVS
Escuela Politécnica Superior
Universidad Autónoma de Madrid, Spain

(Visiting scholar at)
Pattern Recognition and Image Proc. Lab.
Dept. of Computer Science & Engineering
Michigan State University

August 13, 2008 - PRIP Seminar - Dept. Computer Science and Engineering, Michigan State University