# Impact of time variability in off-line writer identification and verification

— Source link ↗

Fernando Alonso-Fernandez, Julian Fierrez, Almudena Gilperez, Javier Ortega-Garcia

**Institutions:** Autonomous University of Madrid

**Topics:** Signature recognition, Handwriting recognition and Biometrics

Related papers:

- A Study on the Development of Embedded Serial Multi-modal Biometrics Recognition System

- A model for comparing signatures

- Automatic person identification system using handwritten signatures

- Online Handwriting and Signature Normalization and Fusion in a Biometric Security Application

- Multi-modal biometric authentication system based on face and signature using legion feature estimation technique

Share this paper: 🅕 🆈 in ✉

**Repositorio Institucional de la Universidad Autónoma de Madrid**

https://repositorio.uam.es

Esta es la **versión de autor** de la comunicación de congreso publicada en:
This is an **author produced version** of a paper published in:

Proceedings of 6th International Symposium on Image and Signal Processing and Analysis, (ISPA) 2009. IEEE, 2009. 540 - 545

**Copyright:** © 2009 IEEE

# Impact of Time Variability in Off-line Writer Identification and Verification

Fernando Alonso-Fernandez, Julian Fierrez, Almudena Gilperez, Javier Ortega-Garcia
Biometric Recognition Group - ATVS - http://atvs.ii.uam.es
Escuela Politecnica Superior - Universidad Autonoma de Madrid
Avda. Francisco Tomas y Valiente, 11 - 28049 Madrid, Spain
{fernando.alonso, julian.fierrez, almudena.gilperez, javier.ortega}@uam.es

## Abstract

*One of the biggest challenges in person recognition using biometric systems is the variability in the acquired data. In this paper, we evaluate the effects of an increasing time lapse between reference and test biometric data consisting of static images of handwritten signatures and texts. We use for our experiments two recognition approaches exploiting information at the global and local levels, and the Biosec-curID database, containing 3,724 signature images and 532 texts of 133 individuals acquired in four acquisition sessions distributed along a 4 months time span. We report results of the recognition systems working both in verification (one-to-one) and identification (one-to-many) mode. The results show the extent of the impact that the time separation between samples under comparison has on the recognition rates, being the local approach more robust to the time lapse than the global one. We also observe in our experiments that recognition based on handwritten texts provides higher accuracy than recognition based on signatures.*

## 1 Introduction

A wide variety of applications require reliable person recognition schemes to either confirm or to determine the identity of an individual. Biometrics refer to the automatic recognition of people based on their physiological or behavioral characteristics [1]. *Physiological biometrics* (e.g. fingerprint, face, iris, etc.) are strong modalities for recognition due to its distinctiveness and reduced subject-specific intra-variability. However, these modalities are usually more invasive and require cooperating subjects. On the other hand, *behavioral biometrics* (e.g. signature, gait, handwritting, keystroking, etc.) are less invasive, but they achieve less recognition accuracy, mainly because lower distinctiveness and larger variability across time.

The problem of writer recognition, which pertains to the category of behavioral biometrics, has received significant interest in recent years. Handwritten signatures as person verification means are widely accepted socially and legally, and are used for that purpose in many transactions daily [2]. On the other hand, the use of handwritten text to identify a person has also received significant interest, mainly due to its application in forensic casework (e.g. crimson notes) [3] and historic document authorship analysis.

There are two main automatic recognition approaches of handwritten material [4]: off-line and on-line. Off-line methods consider uniquely the signature or text image, so only static information is available for the recognition task, which is commonly acquired by document scanning [5]. On the other hand, on-line systems use pen tablets or digitizers which capture dynamic information such as velocity and acceleration of the signing and writing process, providing a richer source of information [6]. On-line recognition systems have traditionally shown to be more reliable as dynamic features are more discriminative between subjects and they are harder to imitate [7]. But in spite of its advantages, there are many cases in which online recognition cannot be used because the handwritten material is collected off-line. This is the case of many government/legal/financial transactions that are performed daily. Also, off-line examination is the common type of criminal casework for forensic experts worldwide [3].

This paper addresses the problem of time separation between acquisitions in automatic person authentication based on scanned images of handwritten signatures and texts. The biometric data acquired from an individual during authentication may be very different from the data that was used to generate the reference model, thereby affecting the comparison. Our goal is to determine to what extent recognition rates are degraded when time between sample acquisitions is increased. For this purpose, we use the Biose-curID database [8], which contains handwritten signatures and texts from 133 subjects acquired in 4 different sessions along a 4 months time span. For our recognition experiments, we use two off-line systems based on global [9], and local [10] image analysis. The two systems are evaluated in both verification and identification mode. In verification mode, a one-to-one comparison between two samples is done, with a decision on whether or not the two samples are from the same person. On the other hand, in identification mode, the system identifies an individual by searching the reference models of all the subjects in the database for a match (one-to-many). As a result, the system returns a ranked list of candidates. Ideally, the first ranked candidate
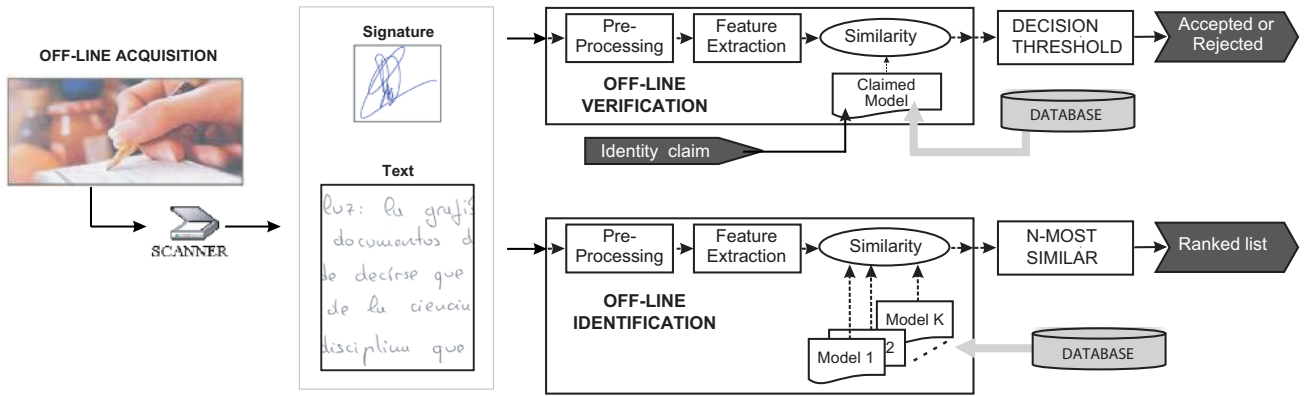
**Figure 1. System model for person verification/identification based on handwritten signature and text images.**

(Top 1) should correspond with the correct identity of the individual, but one can choose to consider a longer list (e.g. Top 10) to increase the chances of finding the correct identity. Identification is a critical component in *negative* recognition applications (or watchlists) where the aim is checking if the person is who he/she (implicitly or explicitly) denies to be, which is a typical situation in forensic/criminal cases [11]. Experiments reported here show the extent of the impact that the time separation between samples being compared has on the recognition rates, both in verification and identification mode. It is also observed in our experiments that using handwritten text images provides higher recognition accuracy than signature images, and that the local system always works better than the global one.

The rest of the paper is organized as follows. The two systems used are described in Section 2. The experimental framework used, including the database and protocol, is described in Section 3. The results obtained are presented in Section 4, and conclusions are finally drawn in Section 5.

## 2 Off-line recognition systems

This section describes the basics of the two recognition systems used in this paper. They exploit information at two different levels. We use an approach based on global analysis, which extracts features from the whole preprocessed image [9], and a second approach based on local image analysis [10]. In Figure 1, the overall model of a verification/identification system is depicted.

### 2.1 Global system

In the global system, input images are first *preprocessed* according to the following consecutive steps (see Table 1): binarization by global thresholding of the histogram [12], and noise removal by morphological closing operation on the binarized image [13]. For the case of signature images, a segmentation of the signature outer traces, and a normalization of the image size to a fixed width of 512 pixels while

**COMMON PREPROCESSING**
- Binarization
- Noise removal

**GLOBAL SYSTEM** (signature only)
- Segmentation
- Size normalization

**LOCAL SYSTEM**
- Component detection
- Contour extraction

**Table 1. Preprocessing stage performed in the global and local systems.**

maintaining the aspect ratio are also carried out. Normalization of signature size is used to make the proportions of different signature realizations of an individual to be the same, whereas segmentation of the outer traces is carried out because a signature boundary typically corresponds to a flourish, which has high intra-user variability [9].

A *feature extraction stage* is then performed, in which slant directions of the strokes and those of the envelopes of various dilated images are extracted using mathematical morphology operators [13], see Figure 2. These descriptors are used as features for recognition as proposed in [14]. For slant direction extraction, the preprocessed image is eroded with 32 structuring elements (EE) like the ones presented in the left column of Figure 2, each one having a different orientation regularly distributed between 0 and 360 degrees [9], thus generating 32 eroded images. A slant direction feature sub-vector of 32 components is then generated, where each component is computed as the signature pixel count in each eroded image. For envelope direction extraction, the preprocessed image is successively dilated 5 times with each one of 6 linear structuring elements, whose orientation is also regularly distributed, thus generating $5 \times 6$ di-

lated images. An envelope direction feature sub-vector of $5 \times 6$ components is then generated, where each component is computed as the signature pixel count in the difference image between successive dilations. The preprocessed signature or text image is finally parameterized as a vector $\mathbf{o} = [o_1, ...o_{62}]$ with 62 components by concatenating the slant and envelope feature sub-vectors. Each client of the system is represented by a statistical model $\boldsymbol{\mu} = [\mu_1, ...\mu_{62}]$ which is estimated by using a reference set of $K$ parameterized images $\{\mathbf{o}_1, ..., \mathbf{o}_K\}$. The parameter $\boldsymbol{\mu}$ denotes the mean vector of the $K$ vectors $\{\mathbf{o}_1, ..., \mathbf{o}_K\}$. In the *similarity computation stage*, to compute the similarity between a claimed model $\boldsymbol{\mu}$ and a parameterized test image $\mathbf{o}$, the $\chi^2$ distance is used:

$$\chi^2_{\mathbf{o}\boldsymbol{\mu}} = \sum_{i=1}^{N} \frac{(o_i - \mu_i)^2}{o_i + \mu_i} \qquad (1)$$

where $N = 62$ is the dimensionality of the vectors $\mathbf{o}$ and $\boldsymbol{\mu}$. Prior to the computation of the $\chi^2$ distance, the vectors $\boldsymbol{\mu}$ and $\mathbf{o}$ are normalized to unit length.

## 2.2 Local system

The *preprocessing stage* of the local system is divided in four parts, as shown in Table 1: binarization by global thresholding of the histogram [12], noise removal by morphological closing operation on the binarized image [13], connected component detection using 8-connectivity, and contour extraction using the Moore's algorithm [13].

In the *feature extraction stage*, curvature of the contour is computed as follows. We consider two contour fragments attached at a common end pixel and compute the directions $\phi_1$ and $\phi_2$ between that pixel and both fragments, see Figure 3. As the algorithm runs over the contour, a joint density function (pdf) $p(\phi_1, \phi_2)$ is then obtained by analyzing in this way the whole processed image, which quantifies the chance of finding two "hinged" contour fragments in the image with angles $\phi_1$ and $\phi_2$, respectively. Each client of the system is represented by a joint pdf that is computed using a reference set of $K$ images. To compute the similarity between a reference model and a given image, the $\chi^2$ distance (Equation 1) is used.

## 3 Database and protocol

We have used for our experiments a sub-corpus of the BiosecurID multimodal database [8], containing handwritten signatures and text from 133 subjects acquired in 4 different sessions distributed along a 4 months time span. Each subject has 4 genuine signatures and 3 forgery signatures per session (from 3 different forgers, the same for the 4 sessions). A Spanish text was also acquired in each session (the same for all subjects and sessions), handwritten in lowercase with no corrections or crossing outs permitted. The resulting sub-corpus has $133 \times 4 \times (4+3) = 3{,}724$ signatures and $133 \times 4 = 532$ texts. All the handwritten data was captured
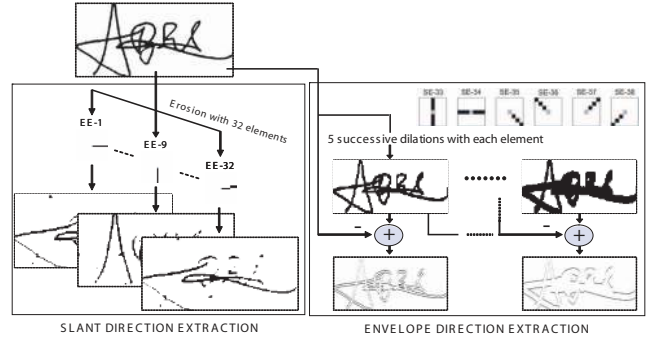


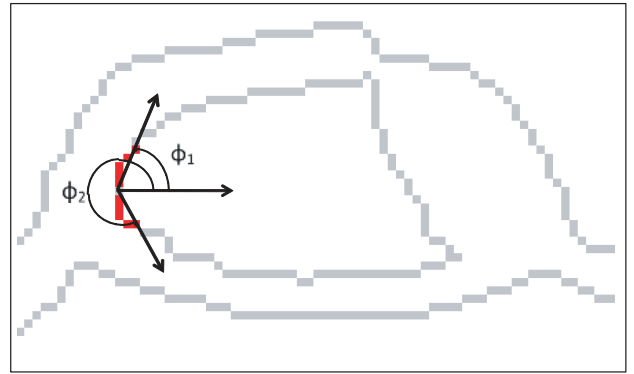**Figure 2. Feature extraction stage performed in the global off-line system.**



**Figure 3. Graphical example of the contour curvature (local off-line system).**

using an inking pen over a Wacom pen tablet so that both on-line dynamic signals and off-line versions (scanned images at 600 dpi) of the data are available. Each signature is written within a $2.5 \times 15$ cm$^2$ frame, and the texts were collected in a different sheet of paper with no guiding lines, just a square frame of $17 \times 16$ cm$^2$ highlighting the writing area. The average amount of text per written sheet is around 9-10 lines in a half A4 page. Some signature and text examples are given in Figure 4. Subjects are modeled for reference using $K=4$ genuine signatures from the first session and $K=1$ page of handwritten text, also from the first session. The remaining signatures and texts are used for testing.

*Verification experiments* with the signature modality are done as follows. Genuine test scores are computed by using the 4 genuine signatures of sessions 2 to 4, and real impostor test scores are computed by using all the available skilled forgeries. As a result, we have $133 \times 4 \times 3 = 1{,}596$ scores from skilled forgeries and three sets of $133 \times 4 = 532$ genuine similarity scores. For the *identification experiments*, we use for testing the 4 genuine signatures of sessions 2 to 4. For each signature, the distances to all the 133 reference models are computed, outputting the $N$ closest identities. An identification is considered successful if the correct identity is
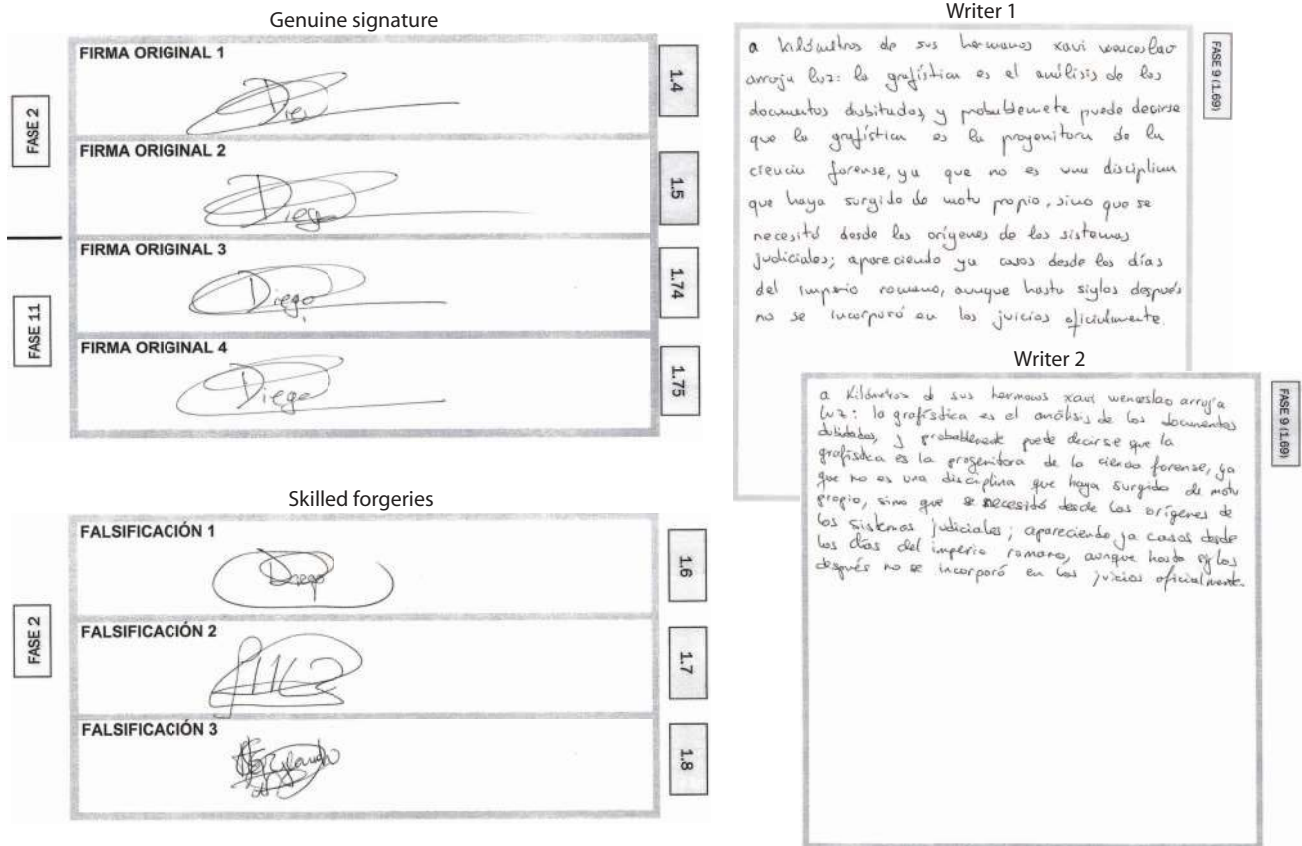
**Figure 4. Signature and text examples from the BiosecurID database [8]. Left: four genuine signatures (top) and three forgeries (bottom). Right: one text example of two different writers.**

among the $N$ outputted ones. As a result, for the identification experiments we have three sets of $133 \times 4 \times 133 = 70{,}756$ similarity scores.

*Verification experiments* with the handwritten texts are as follows. Genuine test scores are computed by using each text page of sessions 2 to 4, and impostor test scores are computed by using all the test pages from the remaining subjects. As a result, we have $133 \times 132 \times 3 = 52{,}668$ scores from impostors and three sets of $133 \times 1 = 133$ genuine similarity scores. For the *identification experiments*, we use the genuine text page of sessions 2 to 4. For each page, the distances to all the reference models are computed, outputting the $N$ closest identities. An identification is considered successful if the correct identity is among the $N$ outputted ones. As a result, we have three sets of $133 \times 133 = 17{,}689$ similarity scores.

## 4 Results

In Figure 5, we show the results for the verification experiments comparing genuine samples from sessions with increasing separation in time. Results are given using either images of handwritten signatures or texts for the same 133 subjects. Verification results in terms of EER (where False Acceptance = False Rejection Rate) are also given in

Figure 7 (left). Similarly, results for the identification experiments are given in Figure 6 and Figure 7 (right).

It is observed from our experiments that the time separation between samples being compared has impact on the recognition rates, both in verification and identification mode. Interestingly enough, we observe however, that once that a minimum time between samples has passed, error rates are not apparently increased. This is observed in Figures 5 and 6, where an small separation between lines marked "Session l vs. Session 3" and "Session l vs. Session 4" can be seen.

Concerning the two modalities evaluated, signature and handwriting, we observe that the latter always provides the highest recognition accuracy. In the verification experiments, the EER using handwritten texts is always below 10% (with an EER of 3% in the best case, see Figure 7). On the contrary, using handwritten signatures, the EER is in the 20-30% range. The explanation is that the texts in our database are written in around half A4 paper sheets, which contain much more discriminative information than signature images, which are done on a $2.5 \times 15$ cm$^2$ frame. Although we are using four signature images for reference, their discriminative information is still much less than the information contained in half page of handwritten text. Similar remarks can be done for the identi-
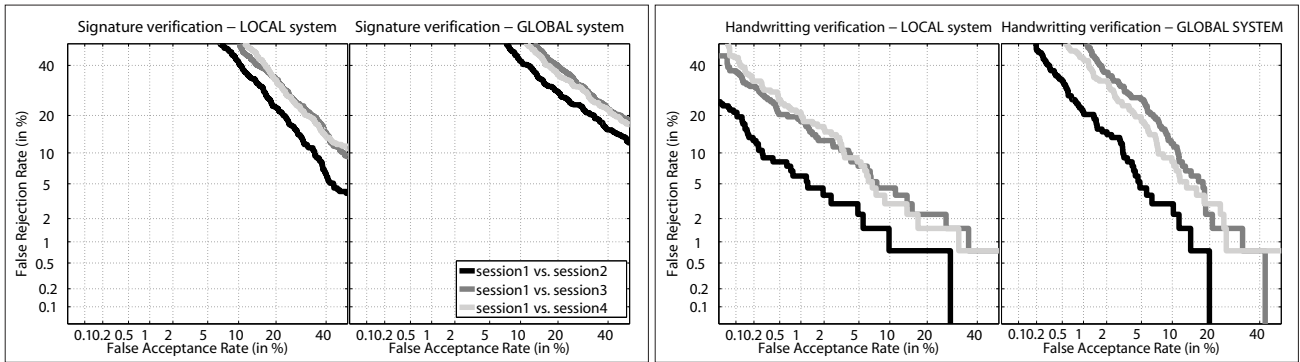
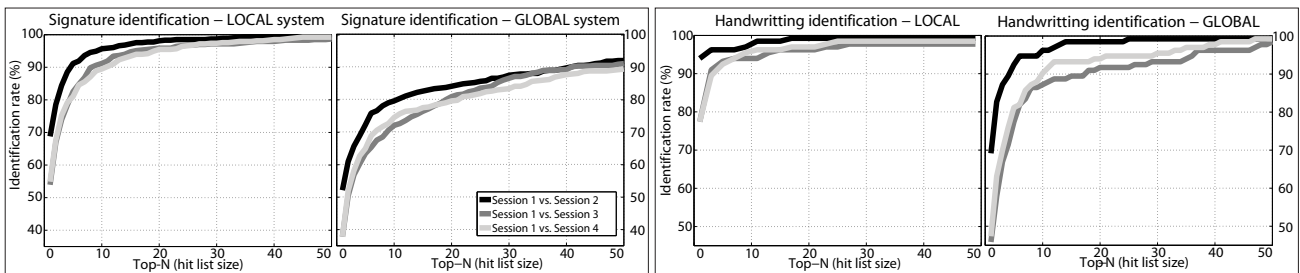**Figure 5. Performance of the verification experiments.**



**Figure 6. Performance of the identification experiments.**

fication experiments. For a hit list size of 10, for instance (see Figure 7), identification rates are mostly above 90% using handwritten texts (with an identification rate of 98.5% in the best case); but using signature images, identification rates are in the 70-90% range in most cases.

Concerning the two recognition algorithms evaluated, we observe from Figures 5 and 6 that the local approach always works better than the global one, either using signatures or texts. This is because the local algorithm processes images locally, thus being able to capture finer details of the image. The global algorithm, on the contrary, processes images as a whole. As a result, it can be seen in Figure 7 that the local approach is less degraded than the global one when time separation between samples is increased (the only exception is the signature verification case). This effect is more evident in the identification case, where the performance of the local approach is only degraded 4.5%, but the global one is degraded 9.5% (when comparing "s1 vs. s2" to "s1 vs. s3").

## 5 Conclusion

This paper has studied the extent of the impact that the time separation between reference and test samples has on the verification and identification of handwritten signatures and text.

Two off-line recognition approaches exploiting information at the global and local levels and the BiosecurID database have been used in our experiments. This database contains scanned signature and text images of 133 individuals acquired in 4 sessions distributed along a 4 months

time span, thus allowing to evaluate time variability. We have carried out experiments both in verification (one-to-one) and identification (one-to-many) mode. We have observed that the time separation between samples being compared has impact on the recognition rates, but once that a specific minimum time between samples has passed (about 2 months), error rates are not apparently worsened with an increased time span between reference and test samples (up to 4 months). This is of course a data-driven statement that should be also studied and validated for longer periods of time (interestingly, new efforts in multimodal database collection have recently enabled this kind of studies for time spans up to a couple of years [15]). The local recognition approach always works better than the global one, both using signatures and texts, and it is less degraded than the global one when time separation between samples is increased. This effect is more evident working in identification mode. We have also observed that recognition based on handwritten text images provides higher accuracy than based on signature images.

Existing technology evaluations have not been aimed to study the effects of time variability in signature and writer recognition [16, 17]. The results of this paper highlight the importance of this phenomenon and encourage its consideration in future technology benchmarks, e.g. [18, 19]. Finally, the results of this paper motivates us to study the individual factors that make some signatures and writers to be more consistent in time than others, in order to develop quality measures that can predict the verification/identification performance [20]. These quality measures can be very useful to compensate the performance
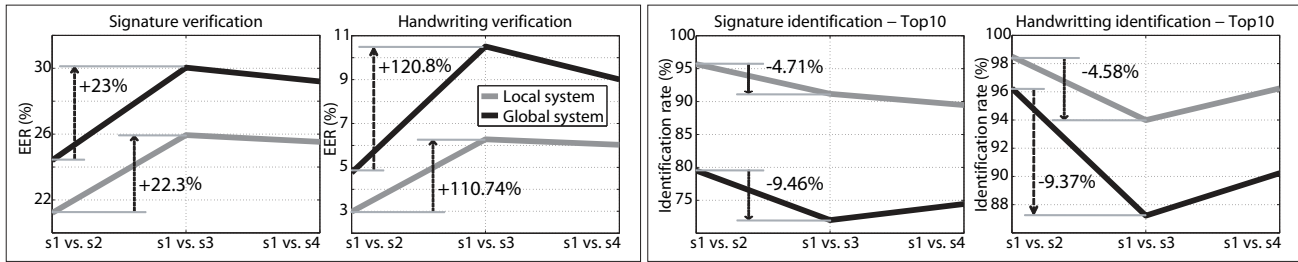
**Figure 7. Verification and identification performance of the signature and handwriting modalities when matching genuine samples from different sessions. Verification results are given in terms of EER (%), while identification experiments are given in terms of success rate (%) for a hit list size of 10. The relative variation of performance is also given. The terms "s1", "s2" and "s3" stand for "session 1", "session 2" and "session 3" respectively.**

drop encountered with increased time spans between reference and test, e.g., using quality-activated template update techniques [21], or quality-based information fusion [22].

## 6  Acknowledgements

## References

[1] A. Jain, A. Ross, S. Pankanti, "Biometrics: A Tool for Information Security", *IEEE Trans. IFS*, 1, 2006, pp. 125–143.

[2] M. Fairhurst, "Signature Verification Revisited: Promoting Practical Exploitation of Biometric Technology", *Electronics & Communication Engineering J.*, 9, 997, pp. 273–280.

[3] S. Srihari, C. Huang, H. Srinivasan, V. Shah, *Digital Document Processing*, ch. 17. Biometric and Forensic Aspects of Digital Document Processing, pp. 379–406. Springer, 2007.

[4] R. Plamondon, S. Srihari, "On-line and Off-line Handwriting Recognition: A Comprehensive Survey", *IEEE Trans. on PAMI*, 22(1), 2000, pp. 63–84.

[5] D. Impedovo and G. Pirlo, "Automatic Signature Verification: The State of the Art", *IEEE Trans. on SMC-C*, 38(5), 2008, pp. 609–635.

[6] J. Fierrez and J. Ortega-Garcia, *Handbook of Biometrics*, chapter 10. On-line Signature Verification, pp. 189–210. Springer, 2008.

[7] G. Rigoll and A. Kosmala, "A Systematic Comparison Between On-line and Off-line Methods for Signature Verification with Hidden Markov Models", *Proc. ICPR*, 2, 1998, pp. 1755–1757.

[8] J. Fierrez, *et al.*, "BiosecurID: A Multimodal Biometric Database", *Pattern Analysis and Applications (accepted)*, 2009.

[9] J. Fierrez-Aguilar, N. Alonso-Hermira, G. Moreno-Marquez, J. Ortega-Garcia, "An Off-line Signature Verification System Based on Fusion of Local and Global Information", *Proc. BIOAW*, Springer LNCS-3087, 2004, pp.295–306.

[10] A. Gilperez, F. Alonso-Fernandez, S. Pecharroman, J. Fierrez, and J. Ortega-Garcia, "Off-line Signature Verification Using Contour Features", *Proc. ICFHR*, 2008.

[11] A. Jain, A. Ross, and S. Prabhakar, "An Introduction to Biometric Recognition", *IEEE Trans. on CSVT*, 14(1), January 2004, pp. 4–20.

[12] N. Otsu, "A Threshold Selection Method for Gray-level Histograms", *IEEE Trans. SMC*, 9, December 1979, pp. 62–66.

[13] R. Gonzalez and R. Woods, *Digital Image Processing*, Addison-Wesley, 2002.

[14] L. Lee and M. Lizarraga, "An Off-line Method for Human Signature Verification", In *Proc. ICPR*, pp. 195–198, 1996.

[15] J. Ortega-Garcia, *et al.*, "The Multi-scenario Multi-environment BioSecure Multimodal Database (BMDB)", *IEEE Trans. on PAMI (to appear)*, 2009.

[16] D. Yeung, H. Chang, Y. Xiong, S. George, R. Kashi, T. Matsumoto, and G. Rigoll, "SVC2004: First International Signature Verification Competition", *Proc. ICBA*, Springer LNCS-3072, July 2004, pp. 15–17.

[17] N. Poh, T. Bourlai, J. Kittler, L. Allano, F. Alonso-Fernandez, O. Ambekar, J. Baker, B. Dorizzi, O. Fatukasi, J. Fierrez, H. Ganster, J. Ortega-Garcia, D. Maurer, A. Salah, T. Scheidat, and C. Vielhauer, "Benchmarking Quality-Dependent and Cost-Sensitive Score-Level Multimodal Biometric Fusion Algorithms", *IEEE Trans. IFS (to appear)*, 2009.

[18] B. Dorizzi, R. Cappelli, M. Ferrara, D. Maio, D. Maltoni, N. Houmani, S. Garcia-Salicetti, and A. Mayoue, "Fingerprint and On-line Signature Verification Competitions at ICB 2009", *Proc. ICB*, LNCS-5558, 2009, pp. 725–732.

[19] SigComp09, "Signature Verification Competition - http://sigcomp09.arsforensica.org", 2009.

[20] F. Alonso-Fernandez, M. Fairhurst, J. Fierrez, J. Ortega-Garcia, "Automatic Measures for Predicting Performance in Off-line Signature", *Proc. ICIP*, 1, September 2007, pp. 369–372.

[21] F. Roli, L. Didaci, and G. Marcialis, "Template Co-update in Multimodal Biometric Systems", *Proc. ICB*, Springer LNCS-4642, 2007, pp. 1194–1202.

[22] J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez, and J. Bigun, "Discriminative Multimodal Biometric Authentication Based on Quality Measures", *Pattern Recognition*, 38(5), 2005, pp. 777–779.