



On-line signature verification system with failure to enrol management

Joan Fàbregas, Marcos Faundez-Zanuy*

Escola Universitària Politècnica de Mataró, Avda. Puig i Cadafalch 101-111, 08303 Mataró, Barcelona, Spain

ARTICLE INFO

Article history:

Received 16 July 2008

Received in revised form 9 December 2008

Accepted 18 January 2009

Keywords:

On-line signatures

Biometric authentication

Feature extraction

Individual threshold

Failure to acquire

Failure to enrol

ABSTRACT

In this paper we introduce a new biometric verification system based on on-line signatures and simulate its operation. For this purpose we have split the MCYT signature database in three subsets: one for classifier training, another for system adjustment and a third one for system testing simulating enrolment and verification. This context corresponds to a real operation, where a new user tries to enrol an existing system and must be automatically guided by the system in order to detect the failure to enrol (FTE) situations. The main contribution of this work is the management of FTE situations by means of a new proposal, called intelligent enrolment, which consists of consistency checking in order to automatically reject low quality samples. This strategy enhances the performance of the system to 22% when 8% of the users are left out. In this situation 8% of the people cannot be enrolled in the system and must be verified by other biometrics or by human abilities. These people are identified with intelligent enrolment and the situation can be thus managed. In addition we also propose a DCT-based feature extractor with threshold coding and discriminability criteria.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Handwritten signatures have a long tradition in commonly encountered verification tasks like, for example, financial transactions and document authentication. They are normally used and well accepted by the general public, and signatures are easily obtained with relatively cheap devices. These are important advantages of signature recognition over other biometrics. Yet, signature recognition has also some drawbacks: it is a difficult pattern recognition problem due to large possible variations between signatures of the same person. These variations may be originated by instabilities, emotions, environmental changes, etc, and are person dependant. In addition, signatures can be forged more easily than other biometrics.

Signature recognition can be split into two categories depending on the data acquisition method:

- *Off-line (static)*: The signature is scanned from a document and the system recognizes the signature analysing its shape.
- *On-line (dynamic)*: The signature is acquired in real time by a digitizing tablet and the system analyses shape and the dynamics of writing, like for example: positions in the x and y axis, pressure applied by the pen, etc.

Using the dynamic data, further information can be inferred, such as accelerations, velocity, curvature radius, etc. [1]. In this paper, we will focus on this approach.

For a signature verification system, depending on testing conditions and environment, three types of forgeries can be established [2]:

- Simple forgery, where the forger makes no attempt to simulate or trace a genuine signature.
- Substitution or random forgery, where the forger uses his/her own signature as a forgery.
- Freehand or skilled forgery, where the forger tries and practices imitating as closely as possible the static and dynamic information of the signature to be forged.

From the point of view of security, the last one is the most damaging; for this reason, some databases suitable for system development include some trained forgeries [3,4].

A comprehensive survey of signature verification can be found in [1,2,5]. The existing methods differ both in feature selection and in decision techniques and they can be divided into two classes [6]:

- *Feature-based*: In which a set of global features is derived from the signature trajectories. Some authors take measures such as total duration of the signature or the number of pen ups, etc. [7,8]. Others compute the Fourier descriptors of the signature

* Corresponding author.

E-mail address: faundez@eupmt.es (M. Faundez-Zanuy).

trajectory [9–11]. Once the features are obtained, the decision method is some kind of distance [12], a statistical model [8], or a neural network [13].

- **Function-based:** In which the time sequences describing the local properties of the signature are used for recognition [14–16]. The most usual ones are position, velocity and acceleration of the trajectory, but also the pressure and the pen inclination [17]. Time sequences are then matched by using elastic distance measures such as dynamic time warping [18–20] or converted into a sequence of vectors and modelled with hidden Markov models [6,21,22].

There are also systems that combine the confidences provided by the two approaches [8,22] and/or use user-dependant decisions, with score normalization techniques [37] or user-dependant thresholds [15], and yield better verification results.

The evaluation of a verification system requires the analysis of two types of errors [24]: Type I error or false reject rate (FRR), which is the percentage of genuine signatures incorrectly rejected by the system. Type II error or false acceptance rate (FAR), which is the percentage of forgeries incorrectly accepted by the system. These two types of errors usually have different associated costs depending on security requirements. The performance of a system may be measured by the value of the detection cost function (DCF) defined as [25]:

$$DCF = C_{FR} \cdot FRR \cdot P(C) + C_{FA} \cdot FAR \cdot P(I) \quad (1)$$

where C_{FR} is the cost of a false reject, C_{FA} is the cost of a false acceptance, $P(C)$ is the prior probability of a client (genuine signature) and $P(I)$ is the prior probability of an impostor (forgery). In this paper, in order to facilitate the comparison with other papers, $C_{FE} = C_{FA} = 1$ and $P(C) = P(I) = \frac{1}{2}$ will be assumed. DCF represents the expected cost of making a detection decision based on a weighted sum of False Rejections and False Acceptance error probabilities. A more meaningful performance measure is the error trade-off curve (DET) [25], which shows how one error changes with respect to the other for several threshold values.

Another measure of a system's performance, often used in biometrics literature, is the equal error rate (EER) [24], which is the point where the FAR and FRR are the same. In order to facilitate comparison with state-of-the-art performances we have also computed the EER where possible. To obtain the EER value, once the tests have been performed, the threshold is trimmed-up in order to balance FAR and FRR. State-of-the-art EER results for skilled forgeries vary around 6% and 3% [4].

In addition to FAR and FRR there are also two other kind of errors that are usually neglected in laboratory conditions, although they are crucial for real world applications [24]. They are named failure to acquire (FTA) and failure to enrol (FTE), and are defined as follows:

Failure to enrol (FTE) rate: FTE rate is the expected proportion of the population for whom the system is unable to generate repeatable templates. This will include those unable to present the required biometric feature, those unable to produce an image of sufficient quality at enrolment, and those who cannot reliably match their template in attempts to confirm that the enrolment is usable.

Failure to acquire (FTA) rate: FTA denotes the proportion of times the biometric device fails to capture a sample when the biometric characteristic is presented to it.

1.1. Motivation of this work

Normally the literature on biometrics deals with the ideal situation where the set of known users is predefined in advance and no new users are added after the initial set up of the system. In some cases, the problem is even more restricted and it is limited to what is known as closed world problems [23]. In this paper, we go

one step further because we deal with the more real situation using two operation steps. First, we adjust a biometric signature recognition system by means of a predefined set of initial users. Second, we study the situation where new users try to enrol in an automatic fashion and the system must be able to detect and manage the FTE situations. That is, the system automatically detects training samples which are unstable to create a consistent model and asks the user for a new biometric sample, we call this procedure 'intelligent enrolment'. This situation is not new in biometrics; ~4% of people cannot be enrolled in a fingerprint system due to the poor quality of their fingerprints. In a signature verification system, due to the large variability among signatures of the same person, this percentage is expected to be bigger. This is a challenging problem that has been neglected in the on-line signature recognition literature.

1.2. Paper organization

This paper is organized as follows: Section 2 provides an overview of the initial system setup. Section 3 explains the system operation. Section 4 presents the experimental results using the MCYT database and Section 5 summarizes the main conclusions.

2. Initial system design

In this section we will discuss the initial set up for the on-line signature verification system, considering a database of genuine and impostor users. Usually, a pattern recognition system consists of two main blocks: feature extraction and classifier. Fig. 1 summarizes this scheme for our on-line signature recognition problem. Given a set of measurements provided by a digitizing tablet, we try to obtain the most relevant features for classification and then a classifier is used to compare with a reference.

2.1. Feature extraction

We use a spectral method [9–11], but replacing the usual Fast Fourier Transform by the one dimensional discrete cosine transform (DCT) to perform a low-frequency filtering of temporal variables. The same approach has been previously used for face recognition [26] with spatial variables. This method characterizes the dynamic and global behaviour of the signature by taking some of the first terms of the DCT which is due to the excellent energy compaction of the DCT and the fact that it is a near optimal substitute for the Karhunen–Loeve transform [29]. It does not require pre-processing (re-sampling and smoothing) as other commonly used methods do [15,27].

Another difference with the usual spectral methods (including [28]) is that, with the help of the classifier (see next section), we perform a threshold coding with discriminability criterion and not the usual zonal or threshold coding with representability criterion, commonly used for image compression [29].

The digitizing tablet provides the horizontal and vertical position of the pen over time in equally spaced intervals: $x(n)$ and $y(n)$, $n = 1, \dots, N_s$. It also provides the pressure $p(n)$, the azimuth angle $\gamma(n)$, and the altitude angle $\phi(n)$, of the pen at the same intervals [3] (see Fig. 2). N_s stands for signature time durations in time samples. We complete these five discrete measures with other two calculated discrete functions: the displacement $\delta s(n) = \sqrt{(\delta x)^2 + (\delta y)^2}$, $n = 1, \dots, N_{s-1}$; where $\delta x = x(n) - x(n-1)$ and $\delta y = y(n) - y(n-1)$ are the coordinate differences between two consecutive points; and the curvature angle $\theta(n) = \alpha(n) - \alpha(n-1)$, $n = 1, \dots, N_{s-2}$; where $\alpha(n) = \arctan(\delta y / \delta x)$, $n = 1, \dots, N_{s-1}$.

The feature extraction performs the DCT of each variable (see Fig. 1) and initially takes the first 30 terms (this number will be

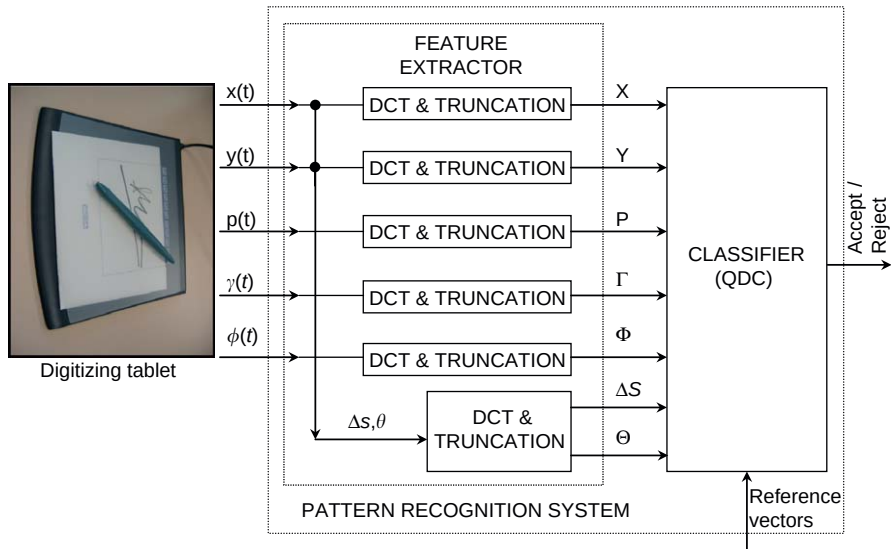


Fig. 1. Pattern recognition system scheme.

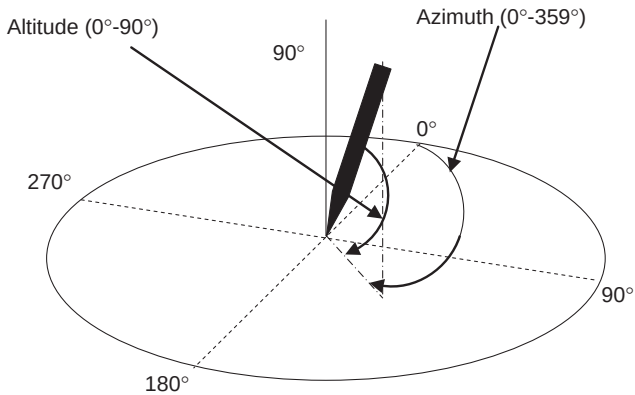


Fig. 2. Azimuth and inclination angles of the pen with respect to the plane of the graphic card.

explained later), representing a total number of 210 features per signature (7×30). Nevertheless, we do not finally pick up all these 30 coefficients, because we will perform a selection using a discriminability criterion, as described in the next section.

2.2. Classifier

In conventional statistical pattern recognition, objects are assigned to classes based on a set of measures extracted from the objects in the form of feature vectors. Usually there are a limited number of classes and a great number of training samples, for example in the recognition of handwritten digits (10 classes) from US postal envelopes described in [30]. This situation is clearly reversed in biometrics, where one normally takes three to five measures per person during the enrolment (three to five samples per class for training) and the enrolled population in the database is large (many more classes than samples per class) [24]. So, the statistical properties of each class are not well defined due to the small number of samples.

In order to overcome this problem, we train a quadratic discriminant classifier (QDC) [31] to solve the dichotomy: *Are these two signatures of the same person?* In doing this, we solve the problem

concerning the number of training samples per class and, because we do not need to train the QDC with the people present in the operational database, it will be capable of classifying in an open world situation [32]. The difference between the signature features by pairs is performed. We compute the covariance matrix Σ_C (C for clients) with pairs of genuine signatures and the covariance matrix Σ_I (I for impostors) with pairs corresponding to a genuine signature and a forgery. Due to the symmetry of the problem (as we can establish the difference between signatures in any order, distributions must be symmetric around the origin) we have $\mu_C = \mu_I = 0$ for the corresponding means. The discriminant function that takes positive values for genuine signature pairs takes the form [31,33]:

$$g(\vec{x}) = \frac{1}{2} \vec{x}^T (\Sigma_I^{-1} - \Sigma_C^{-1}) \vec{x} + \frac{1}{2} \ln \frac{|\Sigma_I|}{|\Sigma_C|} + \ln \frac{P(C)}{P(I)} \quad (2)$$

The discriminative power of each feature may be evaluated comparing the corresponding diagonal terms of Σ_C and Σ_I or their square roots (standard deviations). Let σ_C and σ_I be the standard deviations for clients and impostors of one feature. Deviations between genuine signatures are expected to be less important than the deviations for forgeries. Then, fixing a threshold for the quotient σ_C/σ_I is a good and easy feature selection criteria as we will demonstrate with the experimental results (see Section 4).

Once the features are selected, fixing the σ_C/σ_I threshold, the QDC is trained by computing the covariance matrices, Σ_C and Σ_I , from the training set of measures. By applying the same set to the matcher we can adjust an initial threshold according to a certain criterion, for example a fixed FRR, a fixed FAR or the one that minimizes the DCF. This initial threshold (T_0) is used in the next adjustment stage to determine the operation thresholds of the system.

2.3. System adjustment

In order to adjust the operational thresholds of the system we use another set of measures named validation set. It is composed of genuine signatures and forgeries of persons that have not participated in the training set. First we must decide the number of reference signatures to be used for each person enrolled in the system, usually three to five signatures. After capturing and extracting the features of these signatures we establish their consistency, matching them pair wise. In this article for example, we reject a signature out

of three references if it mismatches the other two. We use the initial threshold, T_0 , obtained in the preceding section. If one signature is rejected it must be substituted by another one from the same person.

We can operate the system with a common threshold (CT) or with individual thresholds (ITs), as described next.

2.3.1. CT operation

Once the enrolment is finished, we use the remaining genuine and forgery signatures to determine the operational thresholds that fulfil a fixed criterion [24], for example a fixed FRR, a fixed FAR or the one that minimizes the DCF (as we have done in this work).

By matching with only one reference, we obtain the operational threshold for enrolment (TE). This threshold is used during the enrolment to compare the captured references pair wise for discarding outliers.

By fusing the scores provided by the references, we can fix the common operational threshold for verification (CT), obtaining the threshold that minimizes the DCF in our case. In this work, we have tested two fusion methods [34]: mean and maximum of the scores. The best results have been obtained with the maximum, as in [15,20].

2.3.2. IT operation

To adapt the verification process to person-dependant variations explained in the first paragraph of the introduction, we have investigated a bi-parametric IT determination based on the adapted decision explained in [35].

Two types of persons are considered:

- *Type 1*: Those with the minimum matching score of the enrolment signatures lying under the CT. For these individuals we expect that lowering the threshold may have a greater effect in reducing the FRR than in increasing the FAR.
- *Type 2*: Those with the minimum matching score of the enrolment signatures lying up to the CT. For these individuals we expect that increasing the threshold may have a greater effect in reducing the FAR than in augmenting the FRR.

To cope with the degree of variation for the ITs the following equation is proposed:

$$IT = (1 - \alpha) \cdot CT + \alpha \cdot \min(ES) \quad (3)$$

where IT is the individual threshold, CT the common threshold, ES the set of enrolment matching scores for the individual and $0 \leq \alpha \leq 1$. The parameters α_1 and α_2 (one for each type of person) must be determined during the system adjustment. We have tested the use of the mean or the total variation instead of the minimum function, but the best results have been obtained by this rule. In this work, after adjusting the CT, already explained, tests are made with different values of the parameters (0, 0.1, 0.2, ..., 1) and we take the ones that produce the best results.

We can also compensate for individual variability with user-dependant score normalization techniques [36,37]. This step is crucial when combining different matchers in a multibiometric system [38], but it is not necessary in our work.

3. System operation

Once the system has been adjusted, it is ready for operation. The system will work with fixed parameters determined by the training and adjustment processes described in Section 2:

- Features selected by the σ_C/σ_I threshold.
- Enrolment threshold, TE.
- Common threshold, CT.
- Parameters that determine the ITs, α_1 and α_2 .

The system operation after system adjustment is described as follows.

3.1. User enrolment

For the enrolment of clients into the system we use the same procedure explained for system adjustment (see the first paragraph of Section 2.3), but with the new enrolment threshold, TE. If one of the reference signatures does not achieve the consistency criteria with the others, the system demands an additional signature sample to the user to replace it. With this procedure we can also help the user deal with the capture process of his/her signature. We can speak of intelligent enrolment because we detect and manage the FTE situations. This is especially important because we avoid the introduction of inconsistent user data which would produce a poor performance of the biometric system.

The people who fail to enrol cannot be managed by the system and we will need to use human abilities or other biometrics to apply with them. This situation is not new in biometrics (for example ~4% of people cannot be enrolled in a fingerprint system [39]) and we can manage this annoyance if we know who they are.

3.2. User verification

In the verification process the user claims his/her identity and the system captures his/her signature. The system matches the captured signature with the references stored during the enrolment of the client (claimed identity), obtaining one score for each reference. The system uses the selected fusion technique and performs an *Accept/Reject* decision comparing the resulting score with the CT or the IT depending on the operation mode.

4. Experiments and results

4.1. Database

We have used our database MCYT [3], acquired with a WACOM graphics tablet. The sampling frequency is set to 100 Hz, obtaining, among other information not used in this paper, the following set for each sampling instant:

- Position in x -axis, $x(t)$: [0–12700], corresponding to 0–127 mm.
- Position in y -axis, $y(t)$: [0–9700], corresponding to 0–97 mm.
- Pressure applied by the pen, $p(t)$: [0–1024].
- Azimuth angle of the pen with respect to the tablet, $\gamma(t)$: [0–3600], corresponding to 0–360°.
- Altitude angle of the pen with respect to the tablet, $\phi(t)$: [300–900], corresponding to 30–90°.

The database consists of 330 people with 25 genuine signatures and 25 skilled forgeries per person. These skilled forgeries are produced by five users after observing the images of the signature to imitate, testing to copy them—almost 10 times—and then, producing the valid forgeries in an easy way (i.e. each individual acting as a forger is requested to sign naturally, without artefacts, such as breaks or slowdowns). We displaced the origin of coordinates to the first point of the signature and normalized spatial distances according to the vertical range of the signature. We do not need a more sophisticated normalization method such as the centre of mass, because we will perform the DCT of the signature and we will look for relative frequencies.

For our experiments users, signatures and forgeries have been randomly reordered in order to mitigate order correlations due to session capture and sub-corpus of the MCYT (see [3]). We split the

Table 1

Features selected from the DCT for each variable and total number of features selected for different values of the σ_c/σ_l threshold.

Threshold	Features selected from the DCT of							Total
	$x(n)$	$y(n)$	$p(n)$	$\gamma(n)$	$\phi(n)$	$\delta s(n)$	$\theta(n)$	
0.35	2	0	1	1	1	0	0	5
0.40	4	2	3	1	1	1	1	13
0.45	7	8	7	1	1	1	2	27
0.50	10	9	9	1	1	4	3	37
0.55	12	11	9	1	1	7	16	57
0.60	15	14	13	1	2	8	27	80
0.65	16	16	18	3	3	12	30	98
0.70	20	16	20	11	20	18	30	135

Results obtained in the training phase of the QDC with DB1.

Table 2

Performance of the QDC, measured with DCF in %, trained on DB1 for the pairs of genuine–genuine and genuine–forgery signatures of the same training set.

σ_c/σ_l Threshold	Min. DCF
0.35	10.77
0.40	9.58
0.45	9.27
0.50	9.28
0.55	10.49
0.60	11.39
0.65	11.24
0.70	11.08

database into three parts:

- DB1: consisting of 80 people and used for training the QDC.
- DB2: consisting of 80 people and used as a validation set for system adjustment.
- DB3: the rest of the database (170 persons) used for testing the system simulating their operation.

The experiments have been performed with three and five references per person, with common and ITs and with intelligent and normal enrolment (without verifying signature consistence). For intelligent enrolment a maximum of 6 (for three references) and 10 (for five references) samples had been used to test the enrolment of a person, if it has not been possible, a FTE is produced. For three references this implies a total of 170×19 genuine tests plus 170×25 impostor tests. For five references we have 170×15 genuine tests plus 170×25 impostor tests.

4.2. Classifier training and feature selection

During the training of the QDC with DB1 the feature selection is also produced. The number of selected features depends on the σ_c/σ_l threshold considered (see Table 1). The features selected from the DCT of each variable are normally taken in natural order, the threshold functions as a parameter defining the number of terms accepted by the low-frequency filter. We can see that 30 terms per variable are sufficient for our purposes.

In Table 2 we see that the performance of the matcher is affected by the number of selected features. Initially the performance increases with increasing σ_c/σ_l thresholds, because the new features provide more information about the signatures. These new features also contribute with certain noise that masks part of the new information, a maximum of performance is achieved and the performance vanishes with increasing thresholds. The performances shown in Table 2 are optimistic due to the fact that they have been measured with the same set used to train the classifier. In the next

Table 3

Adjustment parameters and performance, measured with DCF in %, obtained for the validation set, DB2.

σ_c/σ_l threshold	Three references						Five references					
	TE	CT	α_1	α_2	Min. DCF		TE	CT	α_1	α_2	Min. DCF	
0.40	17.10	20.57	0.2	0.8	8.75		17.16	20.72	0.1	0.9	7.43	
0.45	28.78	34.18	0.1	0.7	7.45		29.16	36.82	0.6	0.9	6.52	
0.50	35.27	44.69	0.1	0.5	7.85		37.75	46.39	0.5	0.8	6.58	
0.55	43.13	58.52	0.6	0.7	8.33		49.35	62.52	0.5	0.8	7.08	

The performance is measured with the individual threshold operation and the maximum fusion method is used.

section we use the validation set, DB2, to obtain better predictions of performance and adjustment thresholds.

Although pen inclination trajectories have shown some discriminative capabilities in other works [13,40], our results are in line with recent research [6] showing that pen inclination information is not as discriminative as the trajectory information in dynamic signature; only two coefficients are present near the threshold of maximum discriminability (~ 0.5). The features extracted from the position and the pressure applied by the pen represent 28 of the 37 total features.

4.3. System adjustment

The system has been adjusted as explained in Section 2.3, with the validation set DB2. We have chosen to optimize the DCF as the criteria for adjusting the thresholds. The maximum fusion method has obtained the best results, with $\sigma_c/\sigma_l = 0.45$ as the best threshold for feature selection. Table 3 shows the adjustment parameters with the minimum DCF obtained for values of the σ_c/σ_l threshold near the best one. It can be observed that the performances for $\sigma_c/\sigma_l = 0.45$ and $\sigma_c/\sigma_l = 0.50$ are very close, the difference is not statistically significant.

Once the system parameters have been obtained we can deal with system operation.

4.4. System operation performance

In this section we have tested the system by using DB3 and simulating operational conditions.

We have enrolled every user of DB3. The consistency of the enrolment signatures is established by matching them pair wise using the threshold TE; for three references a signature is rejected if it mismatches with the other two and, for five references, if it mismatches with three or more. A maximum of six (for three references) and 10 (for five references) samples had been used to test the enrolment of a person; if it has not been possible a FTE is produced.

After enrolment we have tested the performance of the system with the genuine and impostor tests, and using the predefined parameters CT, α_1 and α_2 (a priori results). We have also computed the a posteriori results, computing new values for the parameters in order to minimize the DCF. Comparing both results we can assess the robustness of the system, in the sense that the adjustment made gives good (robust) or poor results during operation.

We must mention that the a priori results presented in this work do not only fix the parameters that define the IT (α_1 and α_2 in our case) as is the case of other works [8,37], but we have also fixed the decision threshold in order to demonstrate the practical applicability to a real operation condition.

For the two types of operation modes the following results have been achieved.

4.4.1. Common threshold

Table 4 shows the performance of the proposed system. We have obtained these results with intelligent enrolment, CT and using the

Table 4

Performance of the system measured with DCF in % for different values of the σ_c/σ_l threshold.

σ_c/σ_l threshold	Three references		Five references	
	DCF a priori	Min. DCF	DCF a priori	Min. DCF
0.35	10.67	10.48	9.78	9.35
0.40	9.01	8.83	7.98	7.85
0.45	7.89	7.86	7.18	7.10
0.50	8.22	7.99	7.02	6.81
0.55	9.55	9.48	8.10	7.99
0.60	11.03	10.85	9.28	9.23
0.65	11.33	10.92	9.50	9.37
0.70	11.80	11.65	11.04	10.44

We have used intelligent enrolment, common threshold and the maximum fusion method. Results obtained with the testing dataset, DB3.

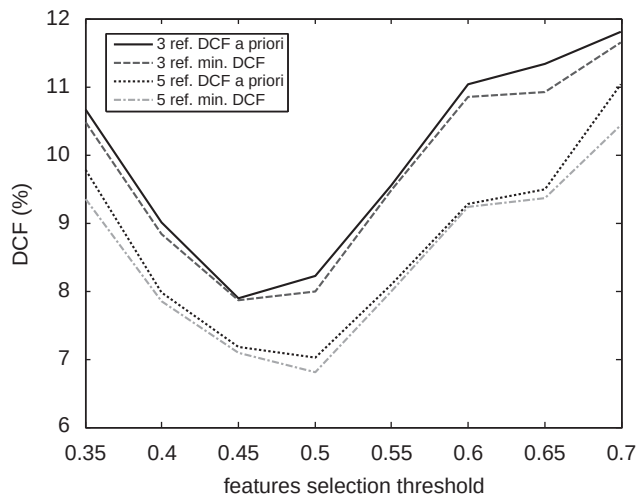


Fig. 3. Plot of the data of Table 4. The feature selection threshold is the σ_c/σ_l threshold and the min. DCF is computed a posteriori.

maximum fusion method. The a priori DCF is calculated using the system in normal operation, with the CT obtained in the adjustment process. The min. DCF is calculated a posteriori, computing a new threshold.

Fig. 3 shows a plot of the performances of the proposed system as they appear in Table 4. The maximum performance (minimum DCF) takes place at the σ_c/σ_l thresholds of 0.45 (three references) and 0.5 (five references). The results a priori and a posteriori are very close, so the proposed system is robust in the sense that the adjustment done gives good results during operation.

Our a posteriori results perform similarly to other state-of-the-art works that use MCYT with five references. They are not directly comparable because the experimental conditions are not the same, but they serve as a reference. With a more complex system like DTW, [20] has obtained a min. DCF of 8.94% and [8], for global information, has obtained an EER of 6.89% (we have computed an EER for five references and $\sigma_c/\sigma_l = 0.50$ of 7.01%).

It is interesting to note that, for five references and $\sigma_c/\sigma_l = 0.50$, the a priori results correspond to a FRR of 7.26% and a FAR of 6.77%, which is quite a balanced situation.

4.4.2. Individual threshold

Table 5 shows the performance of the proposed system. These results have been achieved with intelligent enrolment, ITs and using the maximum fusion method. The a priori DCF is calculated using the system in normal operation, with the IT computed from the CT, α_1 and α_2 obtained in the adjustment process. The min.

Table 5

Performance of the system measured with DCF in % for different values of the σ_c/σ_l threshold.

Threshold	Three references		Five references	
	DCF a priori	Min. DCF	DCF a priori	Min. DCF
0.35	9.27	9.21	8.80	8.00
0.40	7.94	7.12	6.41	6.15
0.45	7.05	6.24	5.26	5.18
0.50	7.16	6.11	5.29	5.17
0.55	6.87	6.75	6.09	5.97
0.60	7.80	7.59	6.73	6.53
0.65	8.25	7.84	8.04	6.90
0.70	8.50	8.53	9.07	7.50

We have used intelligent enrolment, individual threshold and the maximum fusion method. Results obtained with the testing dataset, DB3.

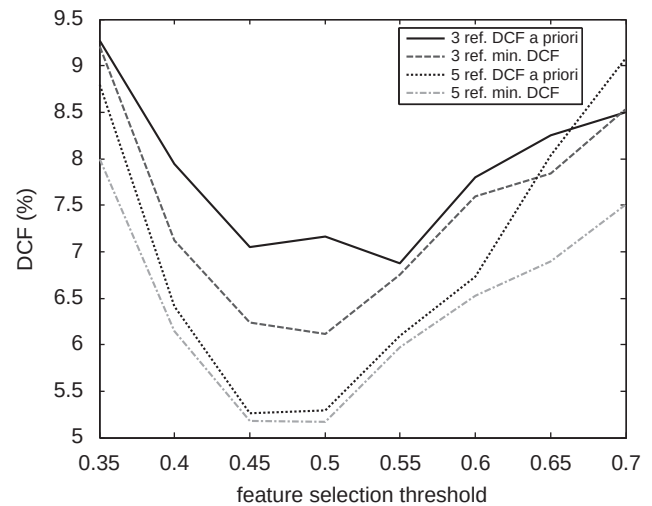


Fig. 4. Plot of the data of Table 5. The feature selection threshold is the σ_c/σ_l threshold and the min. DCF is computed a posteriori.

Table 6

Adjusted with DB2 (a priori) and tested with DB3 (a posteriori) values of the parameters of the system at $\sigma_c/\sigma_l = 0.5$, with five references and using maximum fusion.

	Adjusted (a priori)	Tested (a posteriori)
CT	46.39	48.02
α_1	0.5	0.5
α_2	0.8	0.6

DCF is calculated a posteriori, computing a new CT and news α_1 and α_2 .

Fig. 4 shows a plot of the performances of the proposed system as they appear in Table 5. The maximum performance (minimum DCF) takes place near the σ_c/σ_l thresholds of 0.45 and 0.5. The a priori and a posteriori results near the points of maximum performance are very similar for five references whereas they differ substantially for three references; this is originated by the small number of references in the IT selection. For five references the system proposed is robust in the sense that the adjustment done gives good results during operation.

Compared with the system operating in CT at five references and $\sigma_c/\sigma_l = 0.5$, we achieve an increase of performance of almost 25% in both a priori and a posteriori, cases.

In Table 6 we can compare the values for the adjusted parameters (CT, α_1 and α_2 a priori) with those computed in the test (a posteriori) for $\sigma_c/\sigma_l = 0.5$ and five references. We see that the adjusted

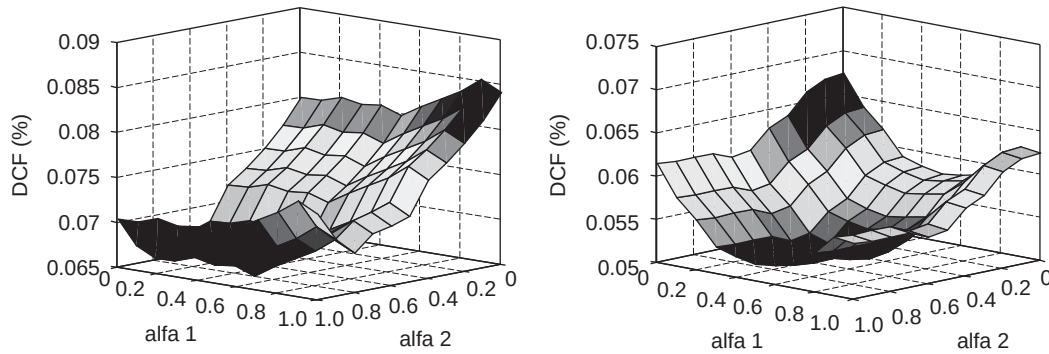


Fig. 5. Left—values of α_1 and α_2 with the corresponding value of the DCF in % ($\sigma_C/\sigma_I = 0.5$ and 5 references) for DB2 (adjustment). The minimum takes place at $\alpha_1 = 0.5$ and $\alpha_2 = 0.8$. Right—the same values for DB3 (test). The minimum takes place at $\alpha_1 = 0.5$ and $\alpha_2 = 0.6$.

Table 7

Results of intelligent enrolment for the feature selection threshold $\sigma_C/\sigma_I = 0.5$.

	Three references		Five references	
	Persons	%	Persons	%
+0	135	79.41	117	68.82
+1	17	10.00	28	16.47
+2	7	4.12	5	2.94
+3	1	0.59	3	1.76
+4	–	–	2	1.18
+5	–	–	1	0.59
FE	10	5.88	14	8.24

The numbers with the plus sign represent the number of extra signatures needed for enrolment. FE is the failure to enrol. Results obtained with the testing dataset, DB3, and the maximum fusion method.

and tested CT are very close because the computed scores spread between -3221.00 and 95.14 . Fig. 5 shows the dependency of the DCF with respect to α_1 and α_2 for DB2 (adjustment) and DB3 (test).

It is interesting to note that for five references and $\sigma_C/\sigma_I = 0.5$ the a priori results correspond to a FRR of 5.77% and a FAR of 4.82%, which is also quite a balanced situation.

4.5. Effects of intelligent enrolment

Table 7 shows the results of intelligent enrolment for the feature selection threshold $\sigma_C/\sigma_I = 0.5$. Remember that we established the consistency of the enrolment signatures by matching them pair wise. For three references we reject a signature if it mismatches with the other two and, for five references, if it mismatches with three or more. A maximum of six (for three references) and 10 (for five references) samples had been used to test the enrolment of a person; if it has not been possible, a FTE is produced. FTE is very important in this system simulation; we expect that in real operation the user will learn, because of system feedbacks, to sign with the digitizing tablet with less difficulty.

In Table 8 we see a comparison of the performances of the system working with intelligent and normal enrolment at $\sigma_C/\sigma_I = 0.5$ with the maximum fusion method. For five references, intelligent enrolment leaves out 8% of the users, and achieves an increase of performance around 22% using common or ITs, in both a priori and a posteriori cases. The IT, as expected, mitigates the negative effects of large variation signers. Intelligent enrolment intends to identify those signers with variation not manageable by the system, and prevents their enrolment. As the same increase in performance for common and IT is obtained, we can see that intelligent enrolment manages part of those people who are not managed by IT. Therefore, IT and intelligent enrolment are complementary.

Table 8

Comparison of the performance of the system ($\sigma_C/\sigma_I = 0.5$ and maximum fusion method) with respect to the type of enrolment (intelligent, normal) and threshold (common, individual).

		Three references		Five references	
		DCF a priori	Min. DCF	DCF a priori	Min. DCF
Intelligent	IT	7.16	6.11	5.29	5.17
	CT	8.22	7.99	7.02	6.81
Normal	IT	7.59	6.59	6.89	6.60
	CT	9.66	9.26	9.16	8.86

The performance is measured with the DCF in %. Results obtained with the testing dataset, DB3.

Fig. 6 shows visual examples of signatures taken from the enrolment process for three references, the first three (top) belong to a user that has been accepted after one attempt only, the second three (middle) to a user that has been accepted with one new attempt (the third signature has been rejected), and the other three (bottom) to a user who has failed to enrol.

We remark that we have worked with a pre-existing database, simulating intelligent enrolment without user interaction. Another advantage of intelligent enrolment that must be proved in a real operating system is the user learning from the system feedback. The feedback of the system will help the user improve their signature abilities with the system interface. These results are encouraging to do future work in this direction.

4.6. Comparison between mean and maximum fusion

The maximum fusion always outperforms the use of the mean fusion in our work, as an example we present in Table 9 the results at $\sigma_C/\sigma_I = 0.5$ for the mean method, for comparison with the same terms of Tables 4 and 5. In Fig. 7 we can compare the DET curves corresponding to the two fusion methods for five references and CT.

4.7. Performance of the system for random forgeries

The system has been trained with skilled forgeries; in this section we study the performance of the system with random forgeries using one authentic signature from all the other enrolled writers as random forgeries. There are a total number of 170×169 random forgeries.

4.7.1. Common threshold

Table 10 shows the performance measured by the DCF expressed in percent. Like in Section 4.2, these results are obtained with



Fig. 6. Visual examples of signatures belonging to users that had been accepted without any new attempt (top), accepted with one new attempt (middle), and not accepted (bottom).

Table 9
Performance of the system ($\sigma_c/\sigma_l = 0.5$) for the mean fusion method with respect to the type of threshold (common, individual).

	Three references		Five references	
	DCF a priori	Min. DCF	DCF a priori	Min. DCF
IT	7.45	7.27	5.77	5.73
CT	9.80	9.20	7.99	7.87

The performance is measured with the DCF in %. Results obtained with the testing dataset, DB3.

intelligent enrolment and using the maximum fusion method. The a priori DCF is calculated by using the system in normal operation, with the CT obtained in the adjustment process. The min. DCF is calculated a posteriori, by computing a new CT.

Fig. 8 shows a plot of performances as they appear in Table 10. The maximum performance (minimum DCF) takes place for σ_c/σ_l near 0.45 and 0.5. The results a priori and a posteriori are not as close as for skilled forgeries; this suggests that training and adjusting the system only with skilled forgeries seems to lower the performance for random forgeries. Nevertheless we still achieve good error values, below 4%. Future work must be done to study how to overcome this effect without lowering the capacity of the system with skilled forgeries.

4.7.2. Individual threshold

Table 11 shows the performance measured by the DCF expressed in %. Like in Section 4.2, these results are obtained with intelligent enrolment and using the maximum fusion method. The a priori DCF is calculated by using the system in normal operation, with the IT computed from the CT, α_1 and α_2 obtained in the adjustment process.

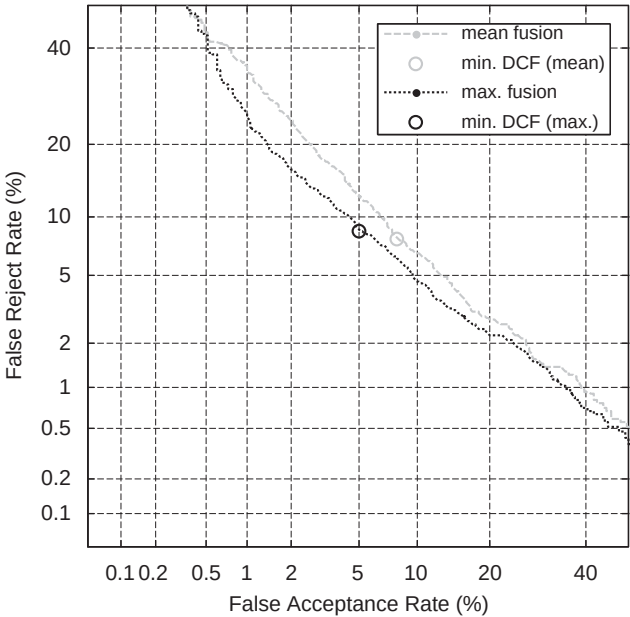


Fig. 7. DET curves for $\sigma_c/\sigma_l = 0.5$ and five references with common threshold. The curves correspond to the mean and maximum fusion methods. Results obtained with the testing data set, DB3.

The min. DCF is calculated a posteriori, computing a new CT and new α_1 and α_2 .

Fig. 9 shows a plot of performances as they appear in Table 11. The maximum performance (minimum DCF) takes place near the

Table 10

Performance of the system for random forgeries measured with DCF in % for different values of the σ_c/σ_l threshold.

Threshold	Three references		Five references	
	DCF a priori	Min. DCF	DCF a priori	Min. DCF
0.35	7.10	5.64	5.63	5.08
0.40	5.06	3.82	3.74	3.11
0.45	3.84	2.71	3.74	2.24
0.50	4.84	2.87	3.74	2.16
0.55	4.83	3.35	4.15	2.61
0.60	5.46	4.08	4.78	3.14
0.65	6.44	4.08	4.95	3.47
0.70	5.53	4.89	6.37	4.12

We have used intelligent enrolment, common threshold and the maximum fusion method. Results obtained with the testing dataset, DB3.

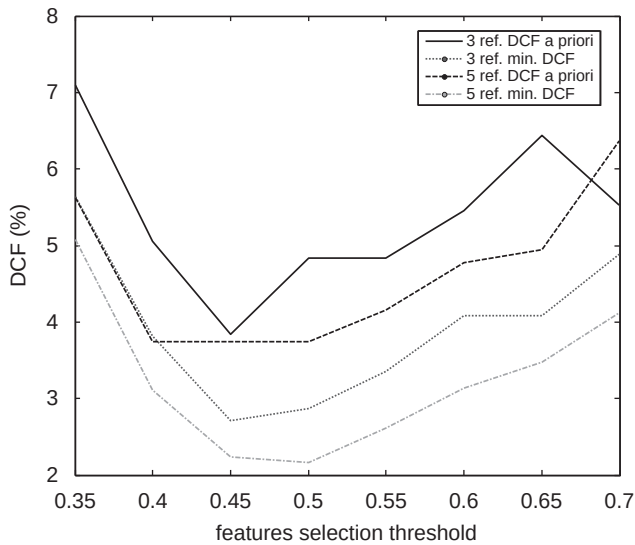


Fig. 8. Plot of the data of Table 10. The feature selection threshold is the σ_c/σ_l threshold and the min. DCF is computed a posteriori.

Table 11

Performance of the system for random forgeries measured with DCF in % for different values of the σ_c/σ_l threshold.

Threshold	Three references		Five references	
	DCF a priori	Min. DCF	DCF a priori	Min. DCF
0.35	6.39	5.41	4.74	4.70
0.40	5.87	3.25	3.45	2.27
0.45	4.97	2.31	2.61	1.83
0.50	5.09	2.40	2.99	1.89
0.55	4.58	2.57	3.27	2.38
0.60	5.34	3.23	3.81	2.69
0.65	5.40	3.44	5.49	2.80
0.70	5.83	4.14	6.13	3.22

We have used intelligent enrolment, individual threshold and the maximum fusion method. Results obtained with the testing dataset, DB3.

σ_c/σ_l thresholds of 0.45 and 0.5. The results a priori and a posteriori near the points of maximum performance are similar for five references whereas they differ substantially for three references; this is caused by the small number of references in the IT selection. For five references the system working with ITs is more robust than for CT.

With the introduction of the IT we achieve an increase of performance of 20% in the a priori case and of 12.5% in the a posteriori case

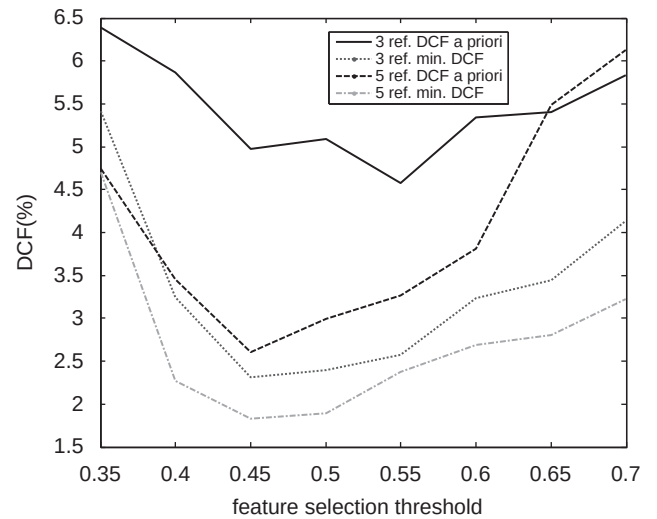


Fig. 9. Plot of the data of Table 11. The feature selection threshold is the σ_c/σ_l threshold and the min. DCF is computed a posteriori.

for five references and $\sigma_c/\sigma_l = 0.5$. These results are in accordance with state-of-the-art methods [4].

5. Conclusions

In this paper, we have presented a new system for on-line signature verification based on DCT feature extraction with discriminability feature selection. We have performed a complete set of simulations with the largest available on-line signature database, MCYT, which consists of 330 people with genuine and skilled forgeries performed by five other different users. The experimentation, although performed in a laboratory, has been developed keeping in mind the casuistic of real world, putting special emphasis on the following problems:

- Open world situation with FTE detection and managing.
- Verification errors obtained with a decision threshold which has been computed a priori.

The main contributions of this work are the following:

- Intelligent enrolment:** A management procedure of the FTE situations by means of a new proposal, “Intelligent enrolment”, which consists of consistency checking in order to automatically reject low quality samples. This strategy enhances the verification errors up to 22%, when 8% of the users are left out, using common or ITs, in both a priori and a posteriori cases. IT and intelligent enrolment are complementary. IT mitigates the negative effects of large variation signers. Intelligent enrolment identifies signers with variation which are not manageable by the system, and prevents their enrolment. In this situation, 8% of the people cannot be enrolled in the system and must be verified by other biometrics or by human abilities. These persons are identified with intelligent enrolment and the situation can thus be managed. This situation is not new in biometrics as ~4% of people cannot be enrolled in a fingerprint system due to the poor quality of their fingerprints. In a signature verification system, due to the large variability among signatures of the same person, this percentage is expected to be bigger. We also expect that some people may be trained by the system feedback during intelligent enrolment to overcome this problem. This work is a

first contribution to this important practical topic in signature verification.

- (2) *Online signature verification system with a new feature extraction:* We have presented a new algorithm based on one dimensional DCT of each vector dimension of feature parameter. Transformed coefficients are selected using a threshold coding with a new discriminability criteria. Experimental results reveal that optimal results are achieved with a total amount of 27 features per signature. This compares favourably with the state-of-the-art algorithms, such as DTW, which requires many more coefficients (several measures for each sampling instant).
- (3) *Bi-parametric adapted schema:* We have proposed a new bi-parametric adapted schema that considers two types of persons for IT calculation. With this approach we achieve an error reduction from 9.26% to 6.59% for three references, and from 8.86% to 6.60% for five references when using normal enrolment.
- (4) *Thresholds set-up a priori:* We have provided verification errors with a priori fixed decision parameters (including the decision threshold), with the system adjusted by means of a validation set. This is another novelty because signature verification literature does not presently include such a priori results, necessary to verify the stability of the system and its practical capabilities. In this most real and worst situation, far from the laboratory condition, we achieve a 5.26% minimum DCF (skilled forgeries) which is comparable to the state-of-the-art performances.

Acknowledgements

This work has been supported by FEDER and MEC, TEC2006-13141-C03-02/TCM

References

- [1] R. Plamondon, S.N. Srihari, On-line and off-line handwriting recognition: a comprehensive survey, *IEEE Trans. PAMI* 22 (1) (2000) 63–84.
- [2] R. Plamondon, G. Lorette, Automatic signature verifications and writer identification—the state of the art, *Pattern Recognition* 1 (2) (1989) 107–131.
- [3] J. Ortega-Garcia, J. Fierrez, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J.J. Igarza, C. Vivaracho, D. Escudero, Q.I. Moro, MCYT baseline corpus: a multimodal biometric database, *IEEE Proc. Vision Image Signal Process.* 150 (2003) 395–401.
- [4] D.Y. Yeung, H. Chang, Y. Xiong, S. George, R. Kashi, T. Matsumoto, G. Rigoll, SVC 2004: First International Signature Verification Competition, *Proceedings of the International Conference on Biometric Authentication*, 2004, in: *Lecture Notes in Computer Science*, vol. 3072, pp. 16–22.
- [5] F. Leclerc, R. Plamondon, Automatic signature verification: the state of the art, *Int. J. Pattern Recognition Artif. Intell.* 8 (3) (1994) 643–660.
- [6] J. Fierrez, J. Ortega-Garcia, D. Ramos, J. Gonzalez-Rodriguez, HMM-based on-line signature verification: feature extraction and signature modeling, *Pattern Recognition Lett.* 28 (2007) 2325–2334.
- [7] H. Ktebgar, J. Richiardi, A. Drygajlo, Global feature selection for on-line signature verification, in: *Proceedings of the 12th International Graphonomics Society Conference*, 2005.
- [8] J. Fierrez-Aguilar, L. Nanni, J. Lopez-Penalba, J. Ortega-Garcia, D. Maltoni, An on-line signature verification system based on fusion of local and global information, audio- and video-based biometric person authentication, 2005, in: *Lecture Notes in Computer Science*, vol. 3546, pp. 523–532.
- [9] C.F. Lam, D. Kamins, Signature recognition through spectral analysis, *Pattern Recognition* 22 (1989) 39–44.
- [10] Q. Wu, S. Lee, I. Jou, On-line signature verification based on logarithmic spectrum, *Pattern Recognition* 31 (1998) 1865–1871.
- [11] Z. Quan, D. Huang, X. Xia, M. Lyu, T. Lok, Spectrum analysis based on windows with variable widths for online signature verification, in: *IEEE 18th International Conference on Pattern Recognition*, 2006.
- [12] L.L. Lee, T. Berger, A. Aviczer, Reliable on-line human signature verification systems, *IEEE Trans. PAMI* 18 (6) (1996) 643–647.
- [13] A. Pascut, A. Czajka, Recognition of human signatures, in: *Proceedings of the International Joint Conference on Neural Networks*, vol. 2, 2001, pp. 1560–1564.
- [14] M.C. Fairhurst, Signature verification revisited: promoting practical exploitation of biometric technology, *IEEE Electron. Commun. Eng. J.* 9 (6) (1997) 273–280.
- [15] A.K. Jain, F.D. Griess, S.D. Connell, On-line signature verification, *Pattern Recognition* 35 (2002) 2963–2972.
- [16] B. Li, D. Zhang, K. Wang, On-line signature verification based on NCA and PCA, *Pattern Anal. Appl.* 8 (2006) 345–356.
- [17] H. Lei, V. Govindaraju, A comparative study on the consistency of features in on-line signature verification, *Pattern Recognition Lett.* 26 (2005) 2483–2489.
- [18] M.E. Munich, P. Perona, Visual identification by signature tracking, *IEEE Trans. PAMI* 25 (2) (2003) 200–217.
- [19] A. Kholmatov, B. Yanikoglu, Identity authentication using improved online signature verification method, *Pattern Recognition Lett.* 26 (2005) 2400–2408.
- [20] M. Faundez-Zanuy, On-line signature recognition based on VQ-DTW, *Pattern Recognition* 40 (2007) 981–992.
- [21] L. Yang, B.K. Widjaja, R. Prasad, Application of hidden Markov models for signature verification, *Pattern Recognition* 28 (2) (1995) 161–170.
- [22] R.S. Kashi, J. Hu, W.L. Nelson, W. Turin, On-line handwritten signature verification using hidden Markov model features, in: *Proceedings of the International Conference on Document Analysis and Recognition*, vol. 1, 1997, pp. 253–257.
- [23] J.L. Wayman, A scientific approach to evaluating biometric systems using mathematical methodology, in: *Proceedings of CardTech/SecureTech*, 1997.
- [24] R.M. Bolle, J.H. Connell, S. Pankanti, N.K. Ratha, A.W. Senior, *Guide to Biometrics*, Springer, Berlin, 2004.
- [25] A. Martin, et al., The DET curve in assessment of detection performance, in: *Proceedings of the Conference on Eurospeech*, vol. 4, 1997, pp. 1895–1898.
- [26] M. Faundez-Zanuy, J. Roure-Alcobe, V. Espinosa-Duro, J.A. Ortega, An efficient face verification method in a transformed domain, *Pattern Recognition Lett.* 28 (2007) 854–858.
- [27] B. Li, D. Zhang, An intelligent online signature verification system, in: *Computer-Aided Intelligent Recognition Techniques and Applications*, Wiley, New York, 2005, pp. 99–117.
- [28] L. Nanni, A. Lumini, A novel local on-line signature verification system, *Pattern Recognition Lett.* 29 (2008) 559–568.
- [29] A.K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1989.
- [30] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning. Data Mining Inference and Prediction*, Springer, Berlin, 2001.
- [31] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, second ed., Wiley-Interscience, New York, 2001.
- [32] J. Fierrez, M. Faundez-Zanuy, Biometric dispersion matcher, *Pattern Recognition* 41 (2008) 3412–3426.
- [33] C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer, Berlin, 2006.
- [34] L.I. Kuncheva, *Combining Pattern Classifiers*, Wiley-Interscience, New York, 2004.
- [35] J. Fierrez-Aguilar, D. Garcia-Romero, J. Ortega-Garcia, J. Gonzalez-Rodriguez, Adapted user-dependent multimodal biometric authentication exploiting general information, *Pattern Recognition Lett.* 26 (16) (2005) 2628–2639.
- [36] A.K. Jain, K. Nandakumar, A. Ross, Score normalization in multimodal biometric systems, *Pattern Recognition* 38 (2005) 2270–2285.
- [37] J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez, Target dependent score normalization techniques and their application to signature verification, *IEEE Trans. Systems Man Cybernet. C* 35 (3) (2005) 418–425.
- [38] A. Ross, K. Nandakumar, A.K. Jain, *Handbook of Multibiometrics*, Springer, Berlin, 2006.
- [39] A.K. Jain, A. Ross, S. Prabhakar, An introduction to biometric recognition, *IEEE Trans. Circuits Systems Video Technol.* 14 (1) (2004).
- [40] D. Sakamoto, H. Morita, T. Ohishi, Y. Komiya, T. Matsumoto, On-line signature verification algorithm incorporating pen position, pen pressure and pen inclination trajectories, in: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2001, pp. 993–996.

About the Author—JOAN FABREGAS was born in Barcelona, Spain in 1961. He received the Ph.D. degree in 2008 from the Valladolid University, and his interest lie in the field of pattern recognition, neural networks and biometrics. He is now Associate Professor at EUP Mataro since 1991.

About the Author—MARCOS FAUNDEZ-ZANUY was born in Barcelona, Spain, in 1969. He received the B.Sc. degree in Telecommunications in 1993 and the Ph.D. degree in 1998, both from the Polytechnical University of Catalunya. He is now Associate Professor at EUP Mataro since 1994 and heads the Signal Processing Group there. His research interests lie in the fields of biometrics and speech coding. He is the Initiator and Chairman of the European COST action 277 “Nonlinear speech processing”. Dr. Faundez-Zanuy is the Spanish Liaison of EURASIP.