



ELSEVIER

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Synthetic on-line signature generation. Part I: Methodology and algorithms

Javier Galbally^{a,*}, Réjean Plamondon^b, Julian Fierrez^a, Javier Ortega-Garcia^a^a Biometric Recognition Group - ATVS, EPS, Universidad Autonoma de Madrid C/ Francisco Tomas y Valiente 11, 28049 Madrid, Spain^b Laboratoire Scribens, Dép. de Génie Électrique, École Polytechnique de Montréal 2900 Boulevard Edouard-Montpetit, Montréal, QC, Canada H3T 1J4

ARTICLE INFO

Article history:

Received 9 September 2010

Received in revised form

28 October 2011

Accepted 10 December 2011

Keywords:

On-line signature

Synthetic generation

Spectral analysis

Kinematic Theory of rapid human

movements

Sigma-Lognormal model

Biometric recognition

ABSTRACT

The theoretical framework and algorithms of a novel method for the generation of synthetic on-line signatures are presented. This model-based approach combines the spectral analysis of real signatures with the Kinematic Theory of rapid human movements in order to generate totally synthetic specimens. Two different algorithms are also described in order to produce duplicated samples from the synthetic master signatures, so that the generation scheme as a whole is able to produce in a complete automatic fashion huge synthetic databases. Typical examples of synthetic specimens are presented to highlight their human-like appearance. The validation protocol and the test results are presented and discussed in a companion paper.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Automatic access of persons to services is becoming increasingly important in the information era. This has resulted in the establishment of a new technological field known as biometric recognition, or simply *biometrics* [1]. The basic aim of biometrics is to discriminate automatically between subjects -in a reliable way and according to some target application- based on one or more signals derived from physical or behavioral traits [2], such as fingerprint [3], face [4], iris [5], voice [6], or written signature [7,8].

One of the big challenges that this relatively new security technology has to face is the permanent need for the collection of new data that permit the objective and statistical evaluation of the performance of biometric recognition systems. In this context, one key element for the development of biometric applications is the availability of biometric databases. In order to comply with this need for new and statistically meaningful data, in recent years important efforts in the form of cooperative national and international projects have been devoted to the acquisition of large multimodal datasets [9–11] (i.e., comprising different biometric traits of the same users). However, the acquisition of biometric features corresponding to a large population of

individuals, together with the desirable presence of biometric variability of each trait (i.e., multi-session, multiple acquisition sensors, different signal quality, etc.), makes database collection a time-consuming, expensive and complicated process, in which a high degree of cooperation of the donors is needed. Additionally, the legal issues regarding data protection are controversial [12,13] and make the sharing and distribution of biometric data among different research groups or industries very tedious and difficult.

Furthermore, these legal restrictions have pushed each research laboratory to acquire their own evaluation data, instead of encouraging the generation of common benchmarks in which to compare the performance of different recognition algorithms in a fair fashion. Only in the frame of technology evaluations the testing of all competing algorithms is carried out on standardized databases and following fixed protocols so that tests are repeatable and results fully comparable. Some examples of these competitive evaluations are the NIST Facial Recognition Technology Evaluations [14]; the NIST Speaker Recognition Evaluations (SRE) [15]; the Fingerprint Verification Competitions (FVC) [16]; the BioSecure Multimodal Evaluation Campaign held in 2007 [17]; or different signature verification competitions [18,19]. However, even in these cases, efforts are punctual and restricted to the duration of the competition, and do not usually remain in time as researchers cannot always access the data to carry their own posterior tests.

In this context, due to the difficulties linked to database acquisition and to the legal obstacles for their free distribution,

* Corresponding author. Tel.: +34 91 497 6210; fax: +34 91 497 2107.

E-mail addresses: javier.galbally@uam.es (J. Galbally), rejean.plamondon@polymtl.ca (R. Plamondon), julian.fierrez@uam.es (J. Fierrez), javier.ortega@uam.es (J. Ortega-Garcia).

in recent years different initiatives have been conducted within the biometric scientific community to generate databases formed by totally synthetic traits [20,21]. These synthetic databases present the advantages of: (i) being effortless to produce (once the generation algorithm has been developed), (ii) having no size restrictions (in terms of subjects and samples per subject) since they are automatically produced from a computer, and (iii) not being subdued to legal aspects because they do not comprise the data of any real user.

Synthetically generated datasets have already been used in some of the previously cited international competitive evaluations for performance assessment tasks [16]. However, their usefulness is not restricted to performance evaluation, and they can also be exploited for other research purposes such as vulnerability assessment (e.g., performing brute force attacks [22], or even inverting certain feature extraction methods [23]), individuality studies in order to better understand the intrinsic information contained within a given trait [24–26], or practical implementations in which the amount of available training data is crucial for decision making [27,28].

However, in spite of their advantages and potential applications, the generation of realistic synthetic biometric data still represents a very complex pattern recognition problem: modeling the information contained in a certain biometric trait as well as the inter-class and intra-class variation found in real databases (i.e., variation between samples of different subjects, and variation between samples of the same subject).

In this work, we address the problem of generating synthetic databases of human-like on-line handwritten signatures. Although the methodology is general, the artificial samples produced follow the pattern of the so called *occidental* signatures which typically consist of left-to-right handwritten concatenated text and some form of flourish (in opposition to other types of signatures consisting of independent symbols such as the *asian* signatures).

The fully automatic approach proposed for the generation of synthetic on-line signatures comprises two successive stages: in the first one, a *master signature* corresponding to a synthetic individual is produced using a generative model based on information obtained from the spectral analysis of real signatures and on the kinematic theory of rapid human movements [29,30] (i.e., this step controls the number of different subjects that will be present in the final synthetic database). In the second step, the master signature is used to generate different samples of that same synthetic subject (i.e., in this second step we generate a number of samples for each user). In the latter stage of the generation scheme, two different novel algorithms for the generation of duplicated samples are proposed: one based on geometric deformations of the signature dynamic functions, and the other on small variations of the Sigma-Lognormal parameters which define each of the strokes forming the master signature [31].

The motivation to base our model on the combined information obtained from the spectral analysis and the Kinematic Theory of rapid human movements comes mainly from three facts:

- Spectral analysis constitutes a general and powerful tool that enables the parameterization of complex time functions such as the ones found in online signature biometrics, and permits to condense the general topological and geometric information shared by real signatures. This is for example patent in [26] where it is used to devise a spectrum-based signature parameterization in order to perform an individuality study of the on-line signature biometrics. Furthermore, working with the spectrum of the signature time functions permits to exploit some similarities that have been observed among different occidental handwritten signatures.

- The Kinematic Theory, which was initially proposed for the analysis of handwriting [29,30] and then used for other applications [32–34], relies on sound mathematical ground to model in a realistic way the different movements involved in handwriting through the application of the Sigma-Lognormal model to parameterize each of the strokes involved in the signing process [33]. Being a theory based on the human writing behavior, its application to the synthetic generation of signatures provides the artificial samples with human-based kinematic information.
- The synthetic generation method, which is based on the previous two complimentary general models, has the advantage of being invariant (in terms of the parameters to be considered, not in their specific values) to cultural or language differences, whereas systems based on visual characteristics often need to be tailored for Chinese, Arabic, European, or American signatures.

The rest of the article is structured as follows. In Section 2 some recent works related to the preset study are given. The overall synthetic generation method is presented in Section 3. The generation of totally synthetic individuals is presented in Section 4, with the two steps (based on spectral analysis and on the Kinematic Theory) involved in the process being described in Sections 4.1 and 4.2. The generation of duplicated samples is presented in Section 5, with the two algorithms proposed for this purpose being described in Sections 5.1 and 5.2. Conclusions are finally drawn in Section 6.

The validation protocol followed to evaluate the proposed approach for the generation of on-line synthetic signatures, together with tests and experimental results are reported in the accompanying paper “Synthetic On-Line Signature Generation. Part II: Experimental Validation”.

2. Related works

Historically, *manually* synthesized biometric traits such as fingerprints and specially signatures and forged handwriting have been a point of concern for experts from a forensic point of view [35,36], and more recently for vulnerability assessment studies [37,38]. However, it has not been until the recent development of the biometric technology when other applications of synthetic samples have been considered and a growing interest has arisen in the scientific community for the analysis of *automatic* generation of synthetic traits such as voice [39], fingerprints [20], iris [21], handwriting [40], face [41], or signature [42].

It should be emphasized that, although there are multiple works which address the problem of generating synthetic traits [43,44], not all of them consider the term *synthetic* in the same way. In particular, three different strategies for producing synthetic biometric samples can be found in the current literature:

- *Duplicated samples*: In this case the generation algorithm starts from one or more *real* samples of a given person and, through different transformations, produces different synthetic (or duplicated) samples corresponding to the same person. This type of algorithms are useful to *increase* the amount of already acquired biometric data but not to generate completely new datasets (i.e., the number of subjects in the final database is restricted to the number of real users available in the original dataset). Therefore, this class of methods can be helpful to synthetically augment the size of the enrollment set of data in identification and verification systems [45–47,28], a critical parameter for instance in signature biometrics [27], but its utility for performance evaluation in biometrics is limited.

The great majority of existing approaches for synthetic signature generation is based on this type of strategy [48–53]. This approach has also been applied to handwriting [54–58], and face synthesis [59,41,60,61].

- **Combination of different real samples:** This is the approach followed by most speech [62,63] and handwriting synthesizers [64–66,34,67,40]. This type of algorithms start from a pool of *real* units, *n*-phones (isolated or combination of sounds) or *n*-grams (isolated or combination of letters), and using some type of concatenation procedure combine them to form the synthetic samples.

Although these techniques are very useful in text-to-speech [68,69], typewriting-to-handwriting [40,70], or CAPTCHAs (Completely Automatic Public Turing Test to tell Computers and Humans Apart) applications [71,72], they present the drawback of needing *real* samples to generate the synthetic trait and therefore their usefulness for performance evaluation in biometrics is also limited (i.e., only samples of the previously acquired real users can be generated). As in the previous case, this perspective for the generation of synthetic data is useful to produce multiple biometric samples of a given real user, but not to generate synthetic individuals and databases (where both control on the number of subjects and samples per subject are needed).

- **Synthetic-individuals:** In this case, some kind of *a priori* knowledge about a certain biometric trait is learned from a development set of real samples (e.g., minutiae distribution, iris structure, signature length, etc.) and then used to create a model that characterizes that biometric trait for a population of subjects. New *synthetic* individuals can then be generated by sampling the constructed model. In a subsequent stage of the algorithm, multiple instances of the synthetic users can be generated by any of the procedures for creating duplicated samples. Regarding performance evaluation and other applications such as vulnerability assessment or individuality studies in biometrics, this approach has the advantage over the two previously presented of not using any real biometric samples in the generation stage to produce completely synthetic databases (i.e., with these strategies there is freedom both in the number of subjects and samples per user to be generated). This way, these algorithms constitute a very effective tool to overcome the usual shortage of biometric data without undertaking highly resource-consuming acquisition campaigns.

Different model-based algorithms have been presented in the literature to generate synthetic individuals for biometric traits such as iris [73,21,74], fingerprint [20], or speech [75,76]. Regarding the signature trait, different methods have been proposed in order to characterize the handwriting process, using the oscillatory motion model [77], the Sigma-Lognormal model [31], or the Beta-Elliptic model [78]. Although all of them have been applied to the analysis and parameterization of signatures and to the generation of duplicated samples, no conclusive experiments have been carried out regarding the suitability of these models for the synthesis of totally artificial subjects. To the best of our knowledge, Popel is the only author who has described this type of approach for synthetic signature generation using a model based on visual characteristics extracted from the time domain [42]. The method was validated visually, comparing the appearance of the synthetic signatures to that of real samples, but no clear quantitative results on the suitability of the technique were given.

3. Synthetic signature generation

The new model-based approach for realistic signature generation proposed in this work is designed to produce samples which coincide to a very high extent with real signatures in terms of: (i) visual appearance, (ii) information content (topological, spectral and kinematic properties), and (iii) performance of the verification systems tested on the synthetic databases. The validation protocol described in Part II of this series of papers gives quantitative results on the *realism* of synthetically produced databases measured according to the previous three characteristics: appearance, information, and performance.

The synthetic generation algorithm, as can be seen in Fig. 1, presents two different stages which will be described in Sections 4 and 5:

- **Stage 1:** A *master signature* corresponding to a synthetic individual is produced using a generative model based both on the spectral and on the kinematical information of real signatures.
- **Stage 2:** In the second stage the master signature is used to generate different samples of that same synthetic user. Two different generation schemes are exploited to produce these duplicated samples. In the sequel, we provide detailed information about this step-wise methodology.

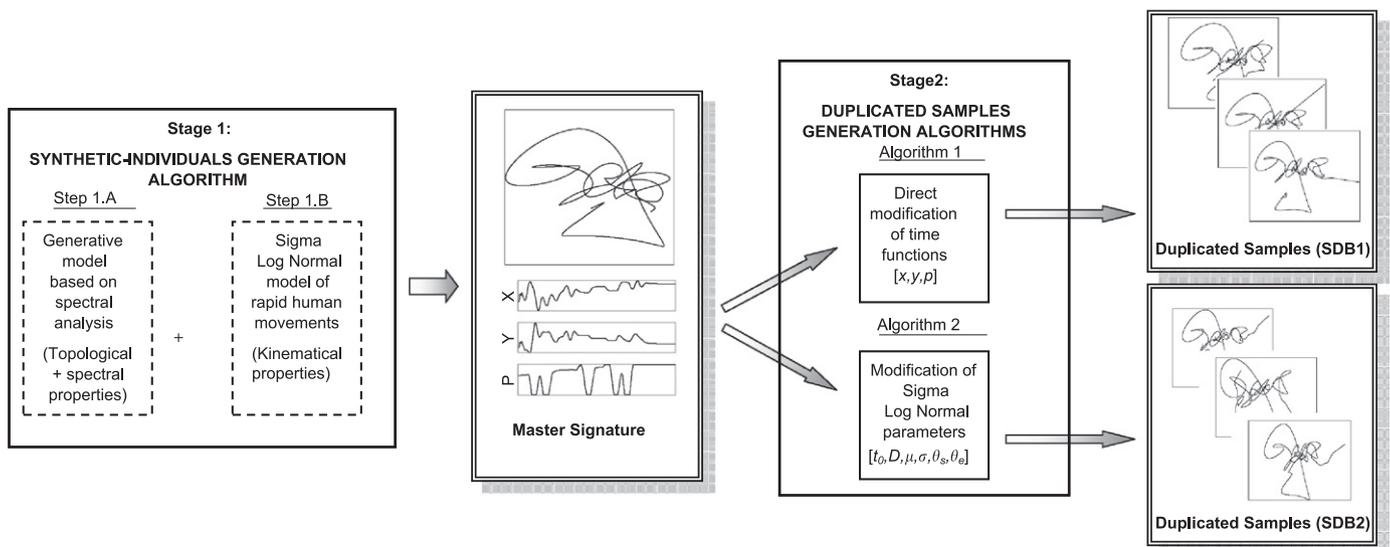


Fig. 1. General architecture of the synthetic signature generation algorithm proposed.

As has been stated before, the proposed generation method may produce different types of signatures (e.g., occidental, asian, arabic) and variability depending on the specific values given to the model parameters. In order to maintain this generality no particular values are given in the description of the generative approach presented in the following sections. All the specific parameter values for a particular case study where the BiosecurID database [10] is used as development set, are given in Part II of this series of articles.

4. Stage 1: generation of synthetic individuals

The objective of this first stage of the proposed synthetic generation method is to produce *one* realistic signature (i.e., *master signature*) of different synthetic individuals, following the inter-variability found in real signatures (i.e., existing variability among signatures produced by different users).

Although other signals such as the azimuth and elevation angles of the input pen might be taken into account, in this work we consider that an online signature is defined by three time sequences $[x[n] \ y[n] \ p[n]]$ specifying respectively the x and y coordinates, and the pressure applied during the signing process at the time instants $n = 1, \dots, N$ (here sampled at 100 Hz).

The algorithm proposed in the present contribution to generate synthetic signers comprises two steps, as can be seen in Fig. 1.

- *Step1.A:* In the first step a parametrical model obtained from the analysis of real signatures in the frequency domain is used to generate a first synthetic master sample with the topological and spectral properties of real samples.
- *Step1.B:* In the second step the synthetic sample produced in the previous step is analyzed and processed according to the Sigma-Lognormal model in order to give the final synthetic master signature the kinematical characteristics of real signatures.

The models involved in both steps are described in the next sections.

4.1. Step 1.A: generative model based on spectral analysis

The algorithm proposed for this first step of the synthetic individuals generation algorithm comprises in turn three successive phases, as can be seen in Fig. 2:

- *Phase 1.A.1:* The synthetic Discrete Fourier Transform (DFT) of the trajectory signals x and y is generated in the frequency domain using a parametrical model obtained by the spectral analysis of a development set of real signatures.

- *Phase 1.A.2:* The resulting trajectory signals are used to generate the pressure function.
- *Phase 1.A.3:* Finally all the three signals are processed in the time domain in order to give the synthetic signatures a more realistic appearance.

4.1.1. Phase 1.A.1: signature model in the frequency domain

The parametrical model proposed in the present contribution is based on the high degree of similarity existing among the trajectory signals of real signatures in the frequency domain. In Fig. 3 some examples of DFTs of the x and y signals are shown, where we can observe that the energy is concentrated in the first coefficients and remains constant and practically negligible from that point (marked with a vertical dashed line in Fig. 3) to the end.

This common structure of the spectrum of x and y allows us to determine a model defined by the following parameters:

- *Sequence length (N):* It defines the number of samples of the three time functions x , y , and p . It is computed for each particular case according to the specific length distribution of the database being used as development set.
- *Number of relevant spectral coefficients (N_R):* It defines the number of coefficients which have a significant power (i.e., those which appear before the dashed line in Fig. 3). This parameter is computed as a percentage of N , $N_R = \lceil \delta_N N \rceil$, where δ_N follows a uniform distribution between $\delta_N^{\min} > 0$ and $\delta_N^{\max} < 1$.
- *Power ratio (G):* Computed as the quotient between the power of the relevant spectral coefficients, and that of the last spectral coefficients (i.e., in Fig. 3 those after the dashed line), $G = P_R/P_l$. The value of G is taken from a uniform distribution, $G \in [G^{\min}, G^{\max}]$.

In order to generate a synthetic signature, the DFT of each of the trajectory signals is generated coloring a white noise sequence of length N_0 with a linear low-pass filter defined by N_{R0} (i.e., filter bandwidth) and G_0 (i.e., attenuation of the high frequencies). This approach implies two simplifications: (i) that all Fourier coefficients are independent (as they are taken from a white noise sequence) and (ii) that both coordinate functions x and y are independent (as they are generated from two white noise sequences).

Although some correlation actually exists among Fourier coefficients corresponding to nearby frequencies, the first of the simplifications (i) has already been applied in other recent studies with fairly good results [26].

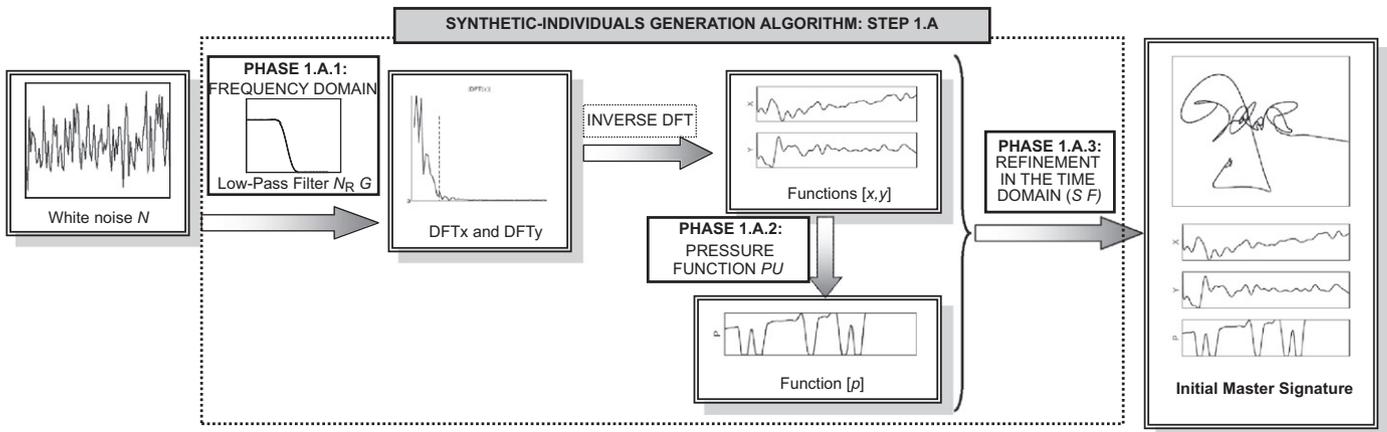


Fig. 2. General diagram of the step 1.A of the synthetic individuals generation algorithm shown in Fig. 1.

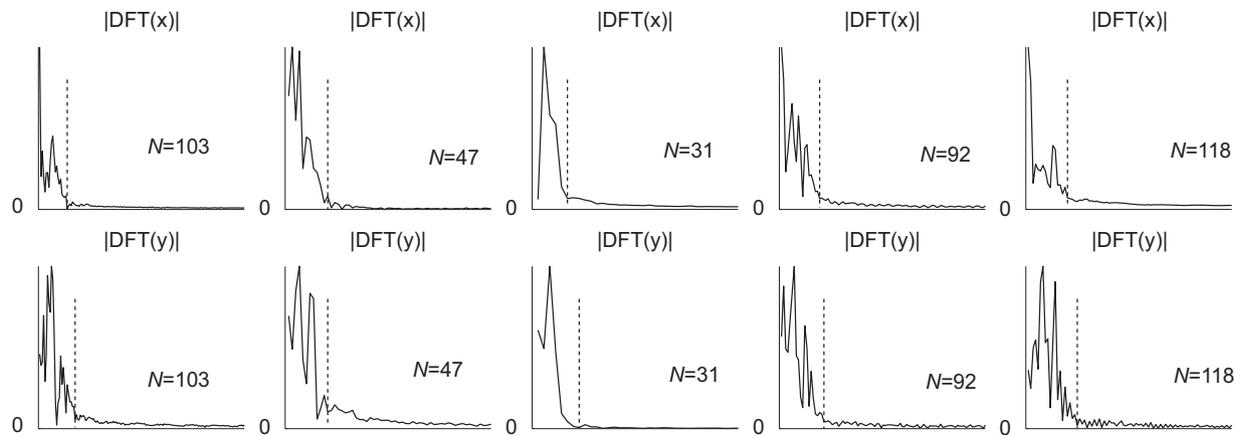


Fig. 3. DFT amplitude examples of the trajectory functions x (top) and y (bottom), of five real signatures (from left to right). The total length N (in number of samples) of the signatures is also given.

As for the second simplification (ii), it could be argued that the coordinate functions x and y are not fully independent. However, this assumption largely helps to keep the simplicity of the method. Furthermore, from the good validation results reported in the second part of this series of two papers [79], it is at least unclear whether the benefits derived from the inclusion of the dependencies between x and y would exceed the inconveniences arisen from a much more complex model.

Once the synthetic DFT of both trajectory signals has been generated, we compute the Inverse DFT (IDFT) in order to obtain the coordinate functions x and y in the time domain.

4.1.2. Phase 1.A.2: the pressure function

The two main features defining the pressure function of a signature are as follows:

- **Number of penups (PU):** A penup is a zero pressure segment of the signature (it occurs when the pen is lifted from the paper during the signing process). The distribution of the number of penups PU is extracted in each case from the specific development database, and applied to the synthetic signatures according to their length N (i.e., a longer signature presents a higher probability of having a large number of penups).
- **Placing of the penups:** From an heuristical analysis of the y and p signals of real signatures, we can conclude that most penups occur close to a singular point of the y function. With this premise, the location of the penups is selected so that they coincide with (or are near) a maximum or minimum of the y signal. Separation between penups is also taken into account at this point in order to avoid placing them unrealistically close. Thus, a minimum distance of n_z non-zero points is set in between consecutive penups. The specific value of this parameter will depend on the sampling rate considered. For a typical sampling rate of 100 Hz assumed in the present research work $n_z=15$.

Once the penups are located through the pressure function, some maximum points (between penups) are selected randomly. In a successive step the pressure waveform is generated by joining all these singular points (penups and maxima) using a cubic spline interpolation algorithm. Once this initial p signal is generated, it is processed in order to avoid undesired effects:

- Many online signature acquisition devices consider 1024 integer pressure levels, so each point of the synthetic p

function is rounded to the nearest integer value, and those which exceed 1024 are set to this maximum value. The same way, those points lower than 0 are set to the penup value.

- A signature pressure signal cannot start or end with a penup. If this is the case the function is artificially changed so that the starting and ending points are non-zero elements.
- Due to the biomechanical properties of the human writing movements, penups cannot be shorter than a certain number of points (around 15 for a 100 Hz sampling rate). The pressure function is accordingly modified in order to avoid unrealistic penups.

4.1.3. Phase 1.A.3: signature refinement in the time domain

Several actions are undertaken at this point to give the signature a more realistic appearance according to occidental samples. This phase of the algorithm may change in order to produce different types of signatures (e.g., arabic or asian).

- **Smoothing:** Both trajectory functions are smoothed using a 10-point moving average in order to avoid possible high frequency noise.
- **Slope (S):** The x function of most left-to-right written signatures presents a general growing tendency fluctuating around a straight of fixed slope. This slope ($S = s_s/s_l$) is artificially produced in this step of the algorithm according to the length of the signature. It is computed as the slope of the diagonal of a rectangle where the long side (s_l) coincides with the length of the signature ($s_l = N$) and the value of the short side (s_s) is extracted as a percentage of the signature length ($s_s = \lceil \delta_s N \rceil$), where δ_s follows a uniform distribution ranging between $\delta_s^{\min} > 0$ and $\delta_s^{\max} < 1$.
- **Flourish (F):** In many cases, real signatures present a large fluctuation of their values at the end of the x and y signals, which in most cases can be identified with a round-like flourish. This final waveform is also artificially added to some signatures in this part of the algorithm (with a probability p_F). It consists of a deformation of the last points of the signature so that its total length is not modified. In order to generate it, (i) its length is fixed as a percentage of the total signature length $F = \delta_F N$ where δ_F follows a uniform distribution between $\delta_F^{\min} > 0$ and $\delta_F^{\max} < 1$, (ii) the maximum and minimum points of the waveform are randomly located and (iii) then they are interpolated using a spline cubic function.
- Additionally, translation, rotation and scaling transformations can also be applied at this point.

4.2. Step 1.B: postprocessing based on the Sigma-Lognormal model

The initial master signature produced in the previous step presents very similar spectral and topological information to that found in real signatures. However, the kinematical properties of this first synthetic sample still differ to some extent to those which are typical of real samples. In particular, very high velocity peaks are observed at the beginning and ending parts of many of these initial master signatures (see dashed circles in Fig. 4), which do not correspond to the typical movement of real signers where the velocity function starts and finishes at zero (or near zero) values [80].

These abnormal velocity artifacts can be corrected by applying the Sigma-Lognormal model to postprocess the initial master signature. This parametrical model is able to represent in a compact manner the kinematical information comprised in humanly produced strokes, therefore it constitutes a useful method to confer the synthetic signatures with the velocity and acceleration properties of the real ones.

This step includes two different phases, as is shown in Fig. 4, where the coordinate functions x and y are slightly modified while the pressure function p remains unaltered:

- **Phase1.B.1:** Extraction of the Sigma-Lognormal parameters. In this phase, the velocity function of the initial synthetic master signature (v_i in Fig. 4) is decomposed in singular strokes and the Lognormal parameters which best fit each of the individual strokes are computed.
- **Phase1.B.2:** Reconstruction of the velocity function of the definitive synthetic master signature according to the previously computed parameters (v_D in Fig. 4). The new coordinate signals are then obtained from the reconstructed velocity function.

4.2.1. Phase 1.B.1: Sigma-Lognormal parameters extraction

The core idea behind our approach relies on the fact that an ideal signature is a well learned movement, executed very fast without any sensorimotor or proprioceptive feedback. Once a subject starts signing, the overall trajectory is executed as planned. This trajectory is made up of individual strokes superimposed in time. Each stroke is characterized by a lognormal velocity profile that reflects the impulse response of the neuromuscular system involved in its production [29]. In the context of the Kinematic Theory and its Sigma-Lognormal model [33], the velocity of the pen tip can be seen as the output of these neuromuscular systems and the signature, as the result of the

vectorial summation of a delayed sequence of N_{LN} strokes:

$$\vec{v}(t) = \sum_{i=1}^{N_{LN}} \vec{v}_i(t; P_i), \quad (1)$$

where each stroke is described by its velocity module:

$$|v_i(t; P_i)| = \frac{D_i}{\sigma(t-t_{0i})\sqrt{2\pi}} \exp\left\{\frac{[\ln(t-t_{0i})-\mu_i]^2}{-2\sigma_i^2}\right\}, \quad (2)$$

and its direction

$$\varphi_i(t; P_i) = \theta_{si} + \frac{\theta_{ei} - \theta_{si}}{2} \left[1 + \operatorname{erf}\left(\frac{\ln(t-t_{0i}) - \mu_i}{\sigma_i\sqrt{2}}\right) \right]. \quad (3)$$

Each curved stroke, indexed by i (with $i = 1 \dots N_{LN}$), is thus completely described in a 2D space by six Sigma-Lognormal parameters $P_i = [D_i, t_{0i}, \mu_i, \sigma_i, \theta_{si}, \theta_{ei}]$:

- D_i : the amplitude of the i th input command.
- t_{0i} : the time occurrence of the input command initiating the i th stroke, a time-shift parameter.
- μ_i : the log-time delay of the i th neuromuscular response expressed on a logarithmic time scale.
- σ_i : the log-response time of the i th neuromuscular response expressed on a logarithmic time scale.
- θ_{si} : starting direction of the i th stroke.
- θ_{ei} : ending direction of the i th stroke.

The velocity components in the Cartesian space can be calculated from the tangential speed as:

$$\vec{v}_x(t) = \sum_{i=1}^{N_{LN}} |v_i(t; P_i)| \cos(\varphi_i(t; P_i)), \quad (4)$$

$$\vec{v}_y(t) = \sum_{i=1}^{N_{LN}} |v_i(t; P_i)| \sin(\varphi_i(t; P_i)). \quad (5)$$

The use of lognormal impulse responses has been shown to reproduce human like movements that encompass all the basic characteristics of the upper limb rapid movements [81]. Moreover, it has been proved recently that such a model constitutes the ultimate kinematic minimization model [82].

The exploitation of the full power of this representation for signature analysis requires the solution of an inverse problem, that is, the recovery of the set of parameters constituting a sequence of strokes. To do so, we have used a software that automatically extracts the parameters that minimize the error

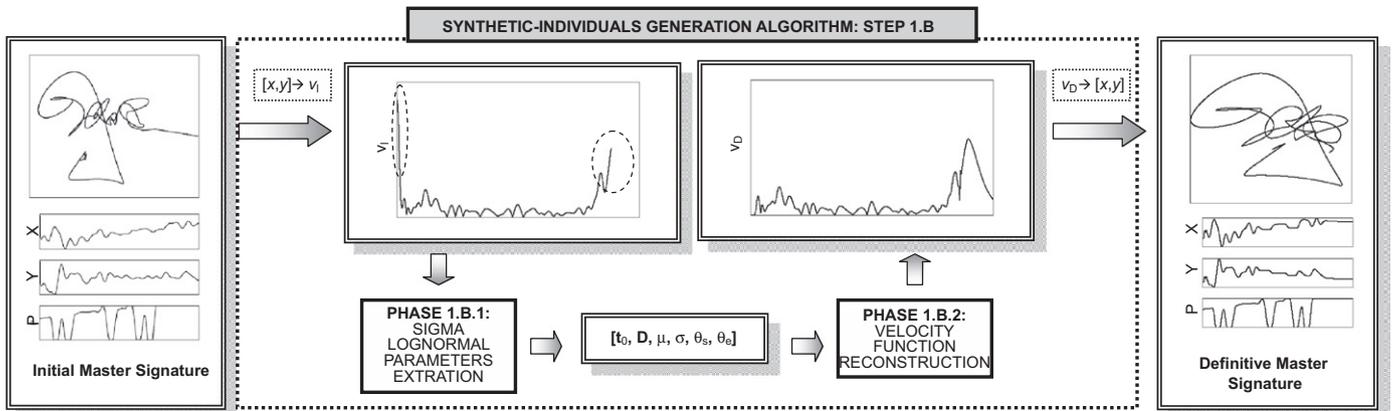


Fig. 4. General diagram of the step 1.B of the synthetic individuals generation algorithm shown in Fig. 1.

between an original signature and its ideal Sigma-Lognormal reconstruction [83]. This extractor works in two different modes:

- In the first, the lognormal equations are estimated and optimized according to their order of occurrence. This mode provides a framework to isolate each lognormal. It is designed such that, while estimating the i th stroke, it minimizes the superposition effects from the direct neighbor strokes (i th-1) and (i th+1) by removing their extracted value. The goodness of the extraction process can be estimated in terms of the Signal to Noise Ratio (SNR) between the original and the reconstructed velocity profiles, defined as:

$$10 \log \left(\frac{\int_{t_s}^{t_e} [v_{x0}^2(t) + v_{y0}^2(t)] dt}{\int_{t_s}^{t_e} [(v_{x0}(t) - v_{xa}(t))^2 + (v_{y0}(t) - v_{ya}(t))^2] dt} \right), \quad (6)$$

where t_s and t_e are respectively the starting and ending times of the signature, and the subindex o refers to the original velocity profile (x or y) while a corresponds to the artificially reconstructed functions.

- If the end of the signals is reached without getting a satisfactory minimal error (as expressed by the SNR), the extractor switches to a second mode where it processes the lognormal strokes in descending order of their area under the curve, that is, according to the importance of their effect on the movement.

The detailed algorithms used by the extraction tool are thoroughly described in [83]. What is of interest for the present study is that such a tool provides, at the end of the process, a list of the parameter values $P = [P_1, P_2, \dots, P_{N_{LN}}]$ that best represent a given target signature. In other words, the algorithm successfully segments a signature into its constituent lognormal strokes.

4.2.2. Phase 1.B.2: reconstruction of the definitive master signature

The set of optimal parameters estimated in the previous phase of the generation algorithm and defining a velocity profile v_D (computed according to Eq. (1)) can then be used as a definitive synthetic master signature from which the signature trajectory function x and y can be reconstructed by:

$$x(t) = \int_0^t |\vec{v}_x(\tau)| d\tau = \sum_{j=1}^{N_{LN}} \frac{D_i}{\theta_{ei} - \theta_{si}} [\sin(\varphi_i(t; P_i)) - \sin(\theta_{si})], \quad (7)$$

$$y(t) = \int_0^t |\vec{v}_y(\tau)| d\tau = \sum_{j=1}^{N_{LN}} \frac{D_i}{\theta_{ei} - \theta_{si}} [-\cos(\varphi_i(t; P_i)) + \cos(\theta_{si})]. \quad (8)$$

In Fig. 5 we show four examples of the initial velocity profiles (v_i) of four synthetic signatures (before step 1.B of the generation algorithm), and their respective definitive reconstructed velocity functions according to the Sigma-Lognormal parameters (v_D , after applying step 1.B), where we can see that the high speed artifacts at the starting and ending segments of the signatures have been corrected, while maintaining the human-like kinematics of the whole trajectory.

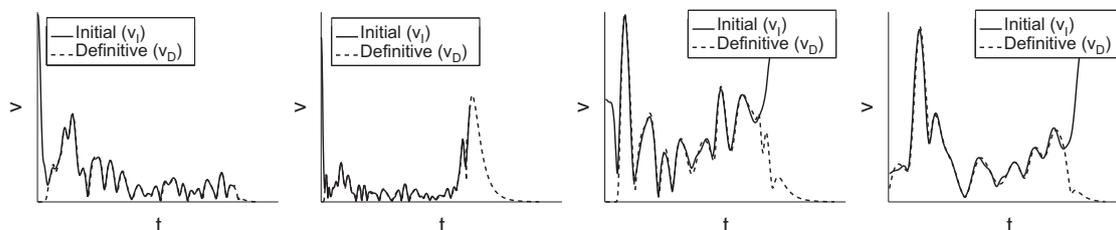


Fig. 5. Initial (v_i) and definitive (v_D) velocity functions of four example synthetic signatures.

5. Stage 2: generation of duplicated samples

Once the time sequences $[x[n] y[n] p[n]]$ defining the master signature of a synthetic user have been generated following the method described in Section 4, the next stage for the automatic generation of synthetic on-line signature databases is the creation of duplicated samples starting from that master sample (as is shown in Fig. 1).

Therefore, the objective of this second stage of the proposed synthetic generation method is to produce different samples of one same synthetic individual following the intra-variability found in real signatures (i.e., existing variability among signatures produced by the same user). For this purpose, two different algorithms are designed.

- *Algorithm 1*: The time sequences of the master signature $[x[n] y[n] p[n]]$ are modified according to a model simulating the distortions introduced by a given channel h . Therefore, this algorithm is based on the direct modification of the spatial and geometry characteristics of the signature.
- *Algorithm 2*: The velocity function v derived from the coordinate functions x and y is decomposed into simple strokes and the Sigma-Lognormal parameters are extracted from each of those individual strokes. Different velocity functions v_j (with $j = 1 \dots S$, being S the number of samples to be generated) are obtained by varying the lognormal parameters, and the corresponding x_j and y_j functions are then recovered from them.

Although it provides very good results (as it will be shown in Part II) and it stands out for its simplicity, Algorithm 1 does not directly rely on biomechanical properties of the handwriting process. The method is based on some signal processing simplifications such as independent noise, and linear distortions. These facts motivated the proposal of Algorithm 2 as an alternative approach for the generation of duplicated samples based on the motion pattern variability rooted in the motor representation space of the handwritten movements. This second algorithm takes advantage of the information extracted using the Sigma-Lognormal model from the neuromuscular impulses related to the movements produced during handwriting in order to take into account effects such as noise correlation and non-linearity of distortions.

The performance of the two algorithms is later evaluated in Part II of this series of articles, where none of them clearly outperforms the other. On the contrary, both methods showed a great degree of complementarity, setting respectively an upper and lower bound to the performance of real samples.

5.1. Algorithm 1: direct modification of the time functions

Lets consider the signing process as follows. A clean dynamic signature $[x[n], y[n], p[n]]$, unique for each subject, is transmitted through an unknown channel h where it is distorted, in this way generating the various genuine impressions corresponding to the natural variability of the subject at hand (see Fig. 6). Under this

framework, the generation of multiple samples from a given clean signature is straightforward once the distortion parameters are set.

In the present algorithm we consider three different phases to model the distortions introduced by the channel h in the signature time signals: (i) adding noise according to a particular Signal to Noise Ratio (SNR), (ii) resampling/downsampling of the original signal by a factor M , and (iii) amplifying or attenuating the signal amplitudes in terms of a parameter α . Next we describe each of the three distortion phases. The specific values of these parameters should be extracted from different development sets of users depending on the type of signatures that want to be produced (e.g., occidental, asian, arabic, etc.)

- **Noise addition (SNR):** Low-frequency noise n_x and n_y is added to the trajectory functions x and y so that the resulting signals x_n and y_n present a particular SNR_x and SNR_y (defined as the quotient between the function's power P_x , and the noise power P_{nx} , i.e., $SNR_x = P_x/P_{nx}$). The SNR should vary depending on whether we want to generate samples from the same or from different sessions (intra- and inter-session SNRs respectively). In our experiments, we assume that the noise is uncorrelated with the signature signals.

At this step, no distortion is introduced in the pressure (p) signal which remains unaltered.

- **Resampling/downsampling (M):** This is equivalent to a duration expansion or contraction of the signals (the same length increase or decrease is applied to all three functions). Considering T as the duration of a signature (the same for the trajectory and pressure signals), the duration of the contracted/expanded new signature is computed as: $T_M = (1 + M)T$. The value of the resampling/downsampling factor M is taken from a different uniform distribution depending on whether we want to produce intra-session ($M \in [-M^{intra}, M^{intra}]$) or inter-session ($M \in [-M^{inter}, M^{inter}]$) variability, being in general $|M^{intra}| < |M^{inter}|$.
- **Amplification/Attenuation (α):** An affine scaling is finally applied to all three signals according to a parameter α (which varies for each time function) [48]. Analogously to the resampling parameter M , the amplification factor α follows a uniform distribution between $[-\alpha_x^{intra}, \alpha_x^{intra}]$ for intra-session samples, and between $[-\alpha_x^{inter}, \alpha_x^{inter}]$ for inter-session samples (similarly for functions y and p). For a given value of the parameter α_x , the scaled function x_α is computed as $x_\alpha = (1 + \alpha_x)x$.

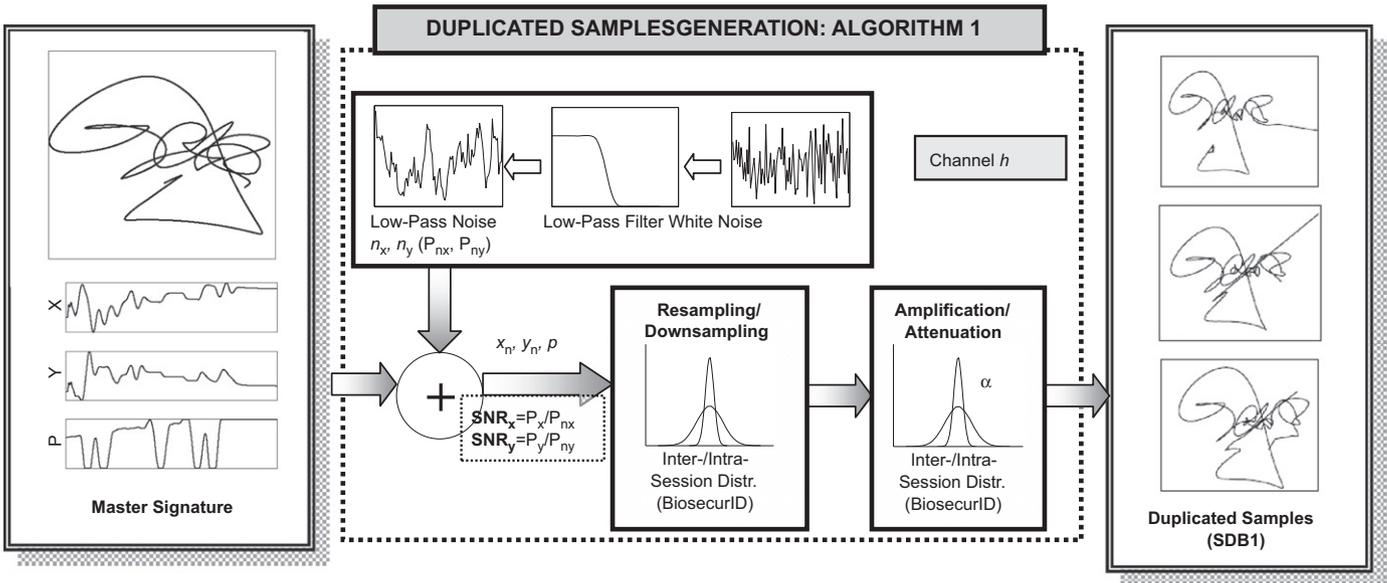


Fig. 6. General architecture of Algorithm 1 for the generation of duplicated samples.

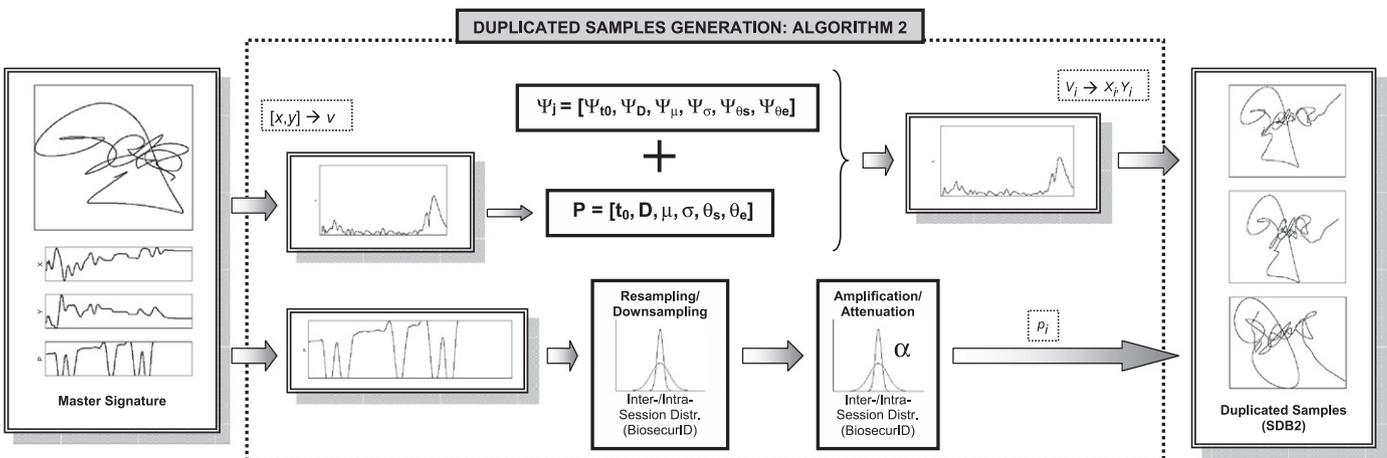


Fig. 7. General architecture of Algorithm 2 for the generation of duplicated samples.

5.2. Algorithm 2: modification of the Sigma-Lognormal parameters

The set of optimal Sigma-Lognormal parameters computed in Section 4.2 can be used as a reference from which a variety of synthetic specimens (duplicated samples) can be generated. Various approaches can be followed here, all based on the same paradigm: adding some noise to the original parameter template, while respecting typical sensitivity patterns previously found in automatic handwriting generation [84].

As described in Section 4.2, the velocity function v of the master signature can be decomposed into single strokes following the Sigma-Lognormal where each stroke s_i (with $i = 1 \dots N_{LN}$, being N_{LN} the total number of strokes in a given signature) is defined by the set of Sigma-Lognormal parameters $P_i = [t_{0i}, D_i, \mu_i, \sigma_i, \theta_{si}, \theta_{ei}]$. Thus, the whole signature is represented by the matrix

$\mathbf{P} = [\mathbf{t}_0, \mathbf{D}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\theta}_s, \boldsymbol{\theta}_e]$, where \mathbf{t}_0 is a column vector of dimension $N_{LN} \times 1$ formed by $[t_{01}, t_{02}, \dots, t_{0N_{LN}}]$ (similar for \mathbf{D} , $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$, $\boldsymbol{\theta}_s$, and $\boldsymbol{\theta}_e$).

The inter-session duplicated samples are generated according to a distortion matrix $\Psi^{\text{inter}} = [\psi_{t_0}^{\text{inter}}, \psi_D^{\text{inter}}, \psi_{\mu}^{\text{inter}}, \psi_{\sigma}^{\text{inter}}, \psi_{\theta_s}^{\text{inter}}, \psi_{\theta_e}^{\text{inter}}]$, where $\psi_{t_0}^{\text{inter}}$ is a column vector of dimension $N_{LN} \times 1$ with its elements belonging to a uniform distribution $[-\psi_{t_0}^{\text{inter}}, \psi_{t_0}^{\text{inter}}]$ (analogously for the rest of distortion vectors comprised in matrix Ψ^{inter}). Then, each of the inter-session samples is computed as $\mathbf{S}_j^{\text{inter}} = \mathbf{P} + \Psi_j^{\text{inter}}$ (with $j = 1 \dots S^{\text{inter}}$, being S^{inter} the number of inter-session samples to be generated).

Therefore, the parameters which define Algorithm 2 for the generation of duplicated samples are the limits of the uniform distribution for each of the lognormal features, i.e., $\Psi = [\psi_{t_0}, \psi_D, \psi_{\mu}, \psi_{\sigma}, \psi_{\theta_s}, \psi_{\theta_e}]$.



Fig. 8. Examples of real (a) and synthetic (b) signatures extracted from MCYT and SDB1, SDB2. Three samples of five different real and synthetic signers are shown together with the time sequences $x[n]$, $y[n]$, and $p[n]$ corresponding to the first sample. (a) Real signatures extracted from the MCYT database [85]. (b) Synthetic signatures produced with the proposed model-based generation algorithm.

Intra-session duplicated samples are generated in a totally analogue way, keeping in general the level of distortion allowed lower than in the inter-session case, $\psi_{t_0}^{\text{inter}} > \psi_{t_0}^{\text{intra}}$ and similarly for the remaining features.

The velocity function v_j is computed from each of the duplicated samples S_j , and in a subsequent step the new coordinate functions x_j and y_j are recovered from the velocity information (according to Eqs. (7) and (8)).

In Algorithm 2 the pressure function p_j of the different duplicated samples is generated following the same process as in Algorithm 1 where it is resampled and amplified according to parameters M and α respectively (see Section 5.1). In this case the resampling parameter M is defined by the length of the new signals x_j and y_j , while α is selected from the same uniform distribution as in Algorithm 1. The general diagram of Algorithm 2 for the synthetic generation of duplicated samples is given in Fig. 7.

As in the previous case, the specific values for the parameters involved in this algorithm and the relationship between the intra- and inter-session features, should be extracted from a development pool of users representative of the type of signatures to be generated.

6. Conclusions

A novel methodology for the generation of synthetic on-line handwritten signatures has been presented. This method combines the advantages of both spectral analysis and the Kinematic Theory of rapid human movements to generate totally synthetic individuals which are not based on any particular real signature. Furthermore, two complimentary algorithms were described for the generation of duplicated samples produced from the artificial specimens, one taking advantage of the benefits of signal processing to directly modify spatial and geometric information and the other exploiting the kinematic properties of rapid human movements. This way, the presented approach as a whole permits the fully automatic generation of huge synthetic on-line signature databases.

As a representative example of the signatures generated with the proposed scheme, three samples of five real (a) and synthetic (b) signers are shown in Fig. 8. This way, we can perform a qualitative (and subjective) validation of the visual appearance of the final synthetic signatures. The real signers are taken from the publicly available MCVT database [85], and the synthetic subjects are produced using the proposed generation method applying Algorithm 1 (first three synthetic individuals) and Algorithm 2 (last two signers) to obtain the duplicated samples. The trajectory and pressure signals of the first signature appear below. We can observe that, although just some recognizable characters can be distinguished in the synthetic signatures, their aspect and that of their time functions is very similar to the real signatures appearance. In Part II of the present work [79], we describe the experimental framework followed to validate from a quantitative and objective perspective the proposed approach, where it is proven that the generation method produces realistic signatures in terms of appearance (for human observers), information contained, and performance (of automatic verification systems).

The novel synthetic generation algorithm described in this work presents a great potential for many different applications such as performance estimation [16], security evaluation in order to test existing biometric solutions against fraudulent access attempts [22], individuality studies [26], or for synthetically increasing the amount of enrollment data in order to improve the performance of a given application [28].

Acknowledgements

J.G. is supported by a FPU fellowship from Spanish MEC. This work has been partially supported by projects Contexts (S2009/TIC-1485) from CAM, Bio-Challenge (TEC2009-11186) from Spanish MICINN, *Dirección General de la Guardia Civil*, and *Cátedra UAM-Telefónica*. R.P. is supported by grant RGPIN-915 from NSERC.

J.G. would also like to thank R.P. as the main coordinator of the Scribens Laboratory at the École Polytechnique de Montréal for hosting him during the development of this research work.

References

- [1] A.K. Jain, A. Ross, S. Pankanti, Biometrics: a tool for information security, *IEEE Transactions on Information Forensics and Security* 1 (2) (2006) 125–143.
- [2] A.K. Jain, A. Ross, P. Flynn (Eds.), *Handbook of Biometrics*, Springer, 2008.
- [3] D. Maltoni, D. Maio, A. Jain, S. Prabhakar, *Handbook of Fingerprint Recognition*, Springer, 2009.
- [4] S.Z. Li, A.K. Jain (Eds.), *Handbook of Face Recognition*, Springer, 2004.
- [5] J. Daugman, How iris recognition works, *IEEE Transactions on Circuits and Systems for Video Technology* 14 (2004) 21–30.
- [6] T. Kinnunen, H. Li, An overview of text-independent speaker recognition: from features to supervectors, *Speech Communication* 52 (2010) 12–40.
- [7] R. Plamondon, S.N. Srihari, On-line and off-line handwriting recognition: A comprehensive survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (1) (2000) 63–84.
- [8] D. Impedovo, G. Piro, Automatic signature verification – state of the art, *IEEE Transactions on Systems, Man and Cybernetics – Part C: Applications and Review* 38 (2008) 609–635.
- [9] J. Fierrez, J. Ortega-García, D. Torre-Toledano, J. Gonzalez-Rodriguez, “BioSec baseline corpus: a multimodal biometric database, *Pattern Recognition* 40 (2007) 1389–1392.
- [10] J. Fierrez, J. Galbally, J. Ortega-García, M.R. Freire, F. Alonso-Fernandez, D. Ramos, D.T. Toledano, J. Gonzalez-Rodriguez, J.A. Siguenza, J. Garrido-Salas, E. Anguiano, G.G. de Rivera, R. Ribalda, M. Faundez-Zanuy, J.A. Ortega, V. Cardeñoso-Payo, A. Viloria, C.E. Vivaracho, Q.I. Moro, J.J. Igarza, J. Sanchez, I. Hernaez, C. Orrite-Uruñuela, F. Martínez-Contreras, J.J. Gracia-Roche, “BioSecurID: a multimodal biometric database, *Pattern Analysis and Applications* 13 (2) (2010) 235–246.
- [11] J. Ortega-García, J. Fierrez, F. Alonso-Fernandez, J. Galbally, M.R. Freire, J. Gonzalez-Rodriguez, C. Garcia-Mateo, J.-L. Alba-Castro, E. Gonzalez-Agulla, E. Otero-Muras, S. Garcia-Salicetti, L. Allano, B. Ly-Van, B. Dorizzi, J. Kittler, T. Bourlai, N. Poh, F. Deravi, M.W.R. Ng, M. Fairhurst, J. Hennebert, A. Humm, M. Tistarelli, L. Brodo, J. Richiardi, A. Drygajlo, H. Ganster, F.M. Sukno, S.-K. Pavani, A. Frangi, L. Akarun, A. Svrnan, The multi-scenario multi-environment BioSecure multimodal database (BMDDB), *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32 (6) (2010) 1097–1111.
- [12] M. Rejman-Greene, Privacy issues in the application of biometrics: a european perspective, *Biometric Systems. Technology, design and performance evaluation*, Springer, 2005., pp. 335–359.
- [13] P.J. Flynn, *Biometric databases*, *Handbook of Biometrics*, Springer, 2007.
- [14] J. Phillips, P. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, in: *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (ICCVPR)*, 2005, pp. 947–954.
- [15] M. Przybocki, A. Martin, NIST speaker recognition evaluation chronicles, in: J. Ortega-García, et al. et al. (Eds.), *ISCA Workshop on Speaker and Language Recognition*, ODYSSEY, 2004, pp. 15–22.
- [16] R. Cappelli, D. Maio, D. Maltoni, J.L. Wayman, A.K. Jain, Performance evaluation of fingerprint verification systems, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (1) (2006) 3–18.
- [17] A. Mayoue, B. Dorizzi, L. Allano, G. Chollet, J. Hennebert, D. Petrovska-Delacretaz, F. Verdet, BioSecure multimodal evaluation campaign 2007 (BMEC 2007), *Guide to Biometric Reference Systems and Performance Evaluation*, Springer, 2009., pp. 327–372.
- [18] D.Y. Yeung, H. Chang, Y. Xiong, S. George, R. Kashi, T. Matsumoto, G. Rigoll, SVC2004: first international signature verification competition, in: D. Zhang, A.K. Jain (Eds.), *Proceedings of the IAPR International Conference on Biometric Authentication (ICBA)*, Springer Lecture Notes in Computer Science, vol. 3072, 2004, pp. 16–22.
- [19] V.L. Blankers, C.E. vand den Heuvel, K.Y. Franke, L.G. Vuurpijl, Icdar 2009 signature verification competition, in: *Proceedings of the International Conference on Document Analysis Recognition ICDAR*, 2009, pp. 1403–1407.
- [20] R. Cappelli, *Synthetic fingerprint generation*, *Handbook of Fingerprint Recognition*, Springer, 2003., pp. 203–231.
- [21] J. Zuo, N.A. Schmid, X. Chen, On generation and analysis of synthetic iris images, *IEEE Transactions on Information Forensics and Security* 2 (2007) 77–90.
- [22] J. Galbally, J. Fierrez, M. Martínez-Díaz, J. Ortega-García, Evaluation of brute-force attack to dynamic signature verification using synthetic samples

- in: Proceedings of the IAPR International Conference on Document Analysis and Machine Intelligence (ICDAR), 2009.
- [23] J. Galbally, R. Cappelli, A. Lumini, G.G. de Rivera, D. Maltoni, J. Fierrez, J. Ortega-Garcia, D. Maio, An evaluation of direct and indirect attacks using fake fingers generated from ISO templates, *Pattern Recognition Letters* 31 (8) (2010) 725–732.
- [24] S. Pankanti, S. Prabhakar, A.K. Jain, On the individuality of fingerprints, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (2002) 1010–1025.
- [25] S.N. Srihari, S.H. Cha, H. Arora, S. Lee, Individuality of handwriting, *Journal of Forensic Science* 47 (2002) 1–17.
- [26] A. Kholmatov, B. Yanikoglu, An individuality model for online signatures using global Fourier descriptors, in: Proceedings of the SPIE Biometric Technology for Human Identification V (BTHI V), vol. 6944, no. 1, 2008.
- [27] J. Fierrez, J. Ortega-Garcia, On-line signature verification, *Handbook of Biometrics*, Springer, 2008., pp. 189–209.
- [28] J. Galbally, J. Fierrez, M. Martinez-Diaz, J. Ortega-Garcia, Improving the enrollment in dynamic signature verification with synthetic samples, in: Proceedings of the IAPR International Conference on Document Analysis and Recognition (ICDAR), 2009.
- [29] R. Plamondon, A kinematic theory of rapid human movements. Part I. Movement representation and generation, *Biological Cybernetics* 72 (1995) 295–307.
- [30] R. Plamondon, A kinematic theory of rapid human movements. Part II. Movement time and control, *Biological Cybernetics* 72 (1995) 309–320.
- [31] M. Djioua, R. Plamondon, A new algorithm and system for the characterization of handwriting strokes with delta-lognormal parameters, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31 (11) (2009) 2060–2072.
- [32] M. Djioua, Contributions a la generalisation, a la comprehension et a l'utilisation de la theorie cinematique dans l'analyse et la synthese du mouvement humain, Ph.D. dissertation, Department of Electrical, Ecole Polytechnique de Montreal, 2007.
- [33] R. Plamondon, M. Djioua, A multi-level representation paradigm for handwriting stroke generation, *Human Movement Science* 25 (2006) 586–607.
- [34] T. Varga, D. Kilchhofer, H. Bunke, Template-based synthetic handwriting generation for the training of recognition systems, in: Proceedings of the Conference of the International Graphonomics Society (IGS), 2005.
- [35] A. Osborn, *Questioned Documents*, Boyd Printing Co., Albany, NY, 1929.
- [36] A. Wehde, J.N. Beffel, *Finger-Prints can be Forged*, Tremonia Publish Co., 1924.
- [37] T. Van der Putte, J. Keuning, Biometrical fingerprint recognition: don't get your fingers burned, in: Proceedings of the Conference on Smart Card Research and Advanced Applications (CARDIS), 2000, pp. 289–303.
- [38] J. Hennebert, R. Loeffel, A. Humm, R. Ingold, A new forgery scenario based on regaining dynamics of signature, in: Proceedings of the IAPR International Conference on Biometrics (ICB), Springer Lecture Notes in Computer Science, vol. 4642, 2007, pp. 366–375.
- [39] T. Dutoit, *An Introduction to Text-to-Speech Synthesis*, Kluwer Academic Publishers, 2001.
- [40] A. Lin, L. Wang, Style-preserving english handwriting synthesis, *Pattern Recognition* 40 (2007) 2097–2109.
- [41] N. Poh, S. Marcel, S. Bengio, Improving face authentication using virtual samples, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2003.
- [42] D.V. Popel, *Signature analysis, verification and synthesis in pervasive environments, Synthesis and Analysis in Biometrics*, World Scientific, 2007, pp. 31–63.
- [43] N.M. Orlans, D.J. Buettner, J. Marques, A survey of synthetic biometrics: capabilities and benefits, in: Proceedings of the International Conference on Artificial Intelligence (ICAI), 2004, pp. 499–505.
- [44] S.N. Yanushkevich, P.S.P. Wang, M.L. Gavrilova, S.N. Srihari (Eds.), *Image Pattern Recognition. Synthesis and Analysis in Biometrics*, World Scientific, 2007.
- [45] V. Margner, M. Pechwitz, Synthetic data for arabic ocr system development, in: Proceedings of the International Conference on Document Analysis and Recognition (ICDAR), 2001.
- [46] T. Varga, H. Bunke, Generation of synthetic training data for an HMM-based handwriting recognition system. in: Proceedings of the International Conference on Document Analysis and Recognition (ICDAR), vol. 1, 2003, pp. 618–622.
- [47] A. Almakour, H. Mouchere, E. Anquetil, Fast online incremental learning with few examples for online handwritten character recognition, in: Proceedings of the International Conference on Frontiers in Handwriting Recognition ICFHR, 2008, pp. 623–628.
- [48] M.E. Munich, P. Perona, Visual identification by signature tracking, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25 (2003) 200–217.
- [49] C. Oliveira, C.A. Kaestner, F. Bortolozzi, R. Sabourin, Generation of signatures by deformations, in: Proceedings of the IAPR International Conference on Advances in Document Image Analysis ICADIA. Springer Lecture Notes in Computer Science, vol. 1339, 1997, pp. 283–298.
- [50] T.J. Hastie, E. Kishon, M. Clark, J. Fan, A Model for Signature Verification, AT&T Bell Laboratories, Technical Report, 1992.
- [51] M. Djioua, C. O'Reilly, R. Plamondon, An interactive trajectory synthesizer to study outlier patterns in handwriting recognition and signature verification, in: Proceedings of the IAPR International Conference on Pattern Recognition (ICPR), 2006, pp. 1124–1127.
- [52] C. Rabasse, R.M. Guest, M.C. Fairhurst, A method for the synthesis of dynamic biometric signature data, in: Proceedings of the IAPR International Conference on Document Analysis and Recognition (ICDAR), 2007.
- [53] J. Richiardi, *Skilled and Synthetic Forgeries in Signature Verification: Large-Scale Experiments*, Institute of Electrical Engineering, Swiss Federal Institute of Technology, Lausanne, Technical Report, 2008.
- [54] S. Setlur, V. Govindaraju, Generating manifold samples from a handwritten word, *Pattern Recognition Letters* 15 (1994) 901–905.
- [55] T.M. Ha, H. Bunke, Off-line, handwritten numeral recognition by perturbation method, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19 (1997) 535–539.
- [56] M. Mori, A. Suzuki, A. Shio, S. Ohtsuka, Generating new samples from handwritten numerals based on point correspondence, in: Proceedings of the IAPR International Workshop on Frontiers in Handwriting Recognition (IWFHR), 2000, pp. 281–290.
- [57] J. Wang, C. Wu, Y.-Q. Xu, H.-Y. Shum, L. Ji, Learning-based cursive handwriting synthesis, in: Proceedings of the IAPR International Workshop on Frontiers of Handwriting Recognition (IWFHR), 2002, pp. 157–162.
- [58] H. Mouchere, S. Bayouhd, E. Anquetil, L. Miclet, Synthetic on-line handwriting generation by distortions and analogy, in: Proceedings of Conference International Graphonomics Society (IGS), 2007, pp. 10–13.
- [59] H.R. Wilson, G. Loffler, F. Wilkinson, Synthetic faces, face cubes, and the geometry of face space, *Vision Research* 42 (2002) 2909–2923.
- [60] H. Wang, L. Zhang, Linear generalization probe samples for face recognition, *Pattern Recognition Letters* 25 (2004) 829–840.
- [61] K. Sumi, C. Liu, T. Matsuyama, Study on synthetic face database for performance evaluation, in: Proceedings of the IAPR International Conference on Biometrics (ICB). Springer Lecture Notes in Computer Science, vol. 3832, 2006, pp. 598–604.
- [62] T. Toda, H. Kawai, K. Shikano, Unit selection algorithm for japanese speech synthesis based on both phoneme and diphone unit, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2002, pp. 465–468.
- [63] A. Black, N. Campbell, Optimizing selection of units from speech database for concatenative synthesis, in: Proceedings of the European Conference on Speech Communication and Technology (EUROSPEECH), 1995, pp. 581–584.
- [64] I. Guyon, Handwriting synthesis from handwritten glyphs, in: Proceedings of the IAPR International Workshop on Frontiers of Handwriting Recognition (IWFHR), 1996, pp. 309–312.
- [65] M. Helmers, H. Bunke, Generation and use of synthetic training data in cursive handwriting recognition, in: Proceedings of Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA). Springer Lecture Notes in Computer Science, vol. 2652, 2003, pp. 336–345.
- [66] J. Wang, C. Wu, Y.-Q. Xu, H.-Y. Shum, Combining shape and physical models for online cursive handwriting synthesis, *International Journal on Document Analysis Recognition* 7 (2004) 219–227.
- [67] L. Ballard, D. Lopresti, F. Monrose, Forgery quality and its implications for behavioral biometric security, *IEEE Transactions on Systems, Man, and Cybernetics* 37 (2007) 1107–1118.
- [68] D.H. Klatt, Review of text-to-speech conversion for english, *Journal of the Acoustical Society of America* 82 (1987) 737–793.
- [69] 2010, Loquendo, Vocal Technology and Services <<http://www.loquendo.com/en/>>.
- [70] 2010, fontGod Corporation <<http://www.fontgod.com/>>.
- [71] A.O. Thomas, A. Rusu, V. Govindaraju, Synthetic handwritten CAPTCHAs, *Pattern Recognition* 42 (2009) 3365–3373.
- [72] 2010, the CAPTCHA Project <<http://www.captcha.net/>>.
- [73] J. Cui, Y. Wang, J. Huang, T. Tan, Z. Sun, An iris image synthesis method based on pca and super-resolution, in: Proceedings of the IAPR International Conference on Pattern Recognition (ICPR), 2004, pp. 471–474.
- [74] S. Shah, A. Ross, Generating synthetic irises by feature agglomeration, in: Proceedings of the IEEE International Conference on Image Processing (ICIP), 2006, pp. 317–320.
- [75] D.H. Klatt, Software for a cascade/parallel formant synthesizer, *Journal Acoustic Society of America* 67 (1980) 971–995.
- [76] N.B. Pinto, D.G. Childers, A.L. Lalwani, Formant speech synthesis: improving production quality, *IEEE Transactions on Acoustics, Speech and Signal Processing* 37 (1989) 1870–1887.
- [77] J.M. Hollerbach, An oscillation theory of handwriting, *Biological Cybernetics* 39 (1981) 139–156.
- [78] H. Bezine, M. Kefi, M. Alimi, On the beta-elliptic model for the control of the human arm movement, *International Journal of Pattern Recognition* 21 (2007) 5–19.
- [79] J. Galbally, J. Fierrez, J. Ortega-Garcia, R.P. lamondon, Synthetic on-line signature generation. PartII: Experimental validation, *Pattern Recognition*, doi:10.1016/j.patcog.2011.12.007, this issue.
- [80] J. Galbally, R. Plamondon, J. Fierrez, C. O'Reilly, M. Martinez-Diaz, J. Ortega-Garcia, Kinematical analysis of synthetic dynamic signatures using the sigma-lognormal model, in: Proceedings of the IAPR International Conference on Frontiers of Handwriting Recognition (ICFHR), 2010.
- [81] R. Plamondon, A. Alimi, Speed/accuracy tradeoffs in target directed movements, *Behavioral and Brain Sciences* 20 (1997) 279–349.
- [82] M. Djioua, R. Plamondon, The limit profile of a rapid movement velocity, *Human Movement Science* 29 (2010) 48–61. doi:10.1016/j.humov.2009.02.007.

- [83] C. O'Reilly, R. Plamondon, Development of a Sigma-Lognormal representation for on-line signatures, *Pattern Recognition* 42 (2009) 3324–3327.
- [84] M. Djoua, R. Plamondon, Studying the variability of handwriting patterns using the kinematic theory, *Human Movement Science* 28 (2009) 588–601.
- [85] J. Ortega-Garcia, J. Fierrez, D. Simon, M.F. Gonzalez, V. Espinosa, A. Satue, I. Hernaez, J.J. Igarza, C. Vivaracho, D. Escudero, Q.I. Moro, MCYT baseline corpus: a bimodal biometric database, *IEE Proceedings of Vision, Image and Signal Processing* 150 (6) (2003) 391–401.

Javier Galbally received the M.Sc. in Electrical Engineering in 2005 from the Universidad de Cantabria, and the Ph.D. degree in Electrical Engineering in 2009, from Universidad Autonoma de Madrid, Spain. Since 2006 he is with Universidad Autonoma de Madrid, where he is currently working as an assistant researcher. He has carried out different research internships in worldwide leading groups in biometric recognition such as BioLab from Universita di Bologna Italy, IDIAP Research Institute in Switzerland, or the Scribens Laboratory at the École Polytechnique de Montréal in Canada. His research interests are mainly focused on the security evaluation of biometric systems, but also include pattern and biometric recognition, and synthetic generation of biometric traits. He is actively involved in European projects focused on vulnerability assessment of biometrics (e.g. STREP Tabula Rasa) and is the recipient of a number of distinctions, including: IBM Best Student Paper Award at ICPR 2008, finalist of the EBF European Biometric Research Award 2009, and best Ph.D. Thesis Award by the Universidad Autonoma de Madrid 2010.

Réjean Plamondon received a B.Sc. degree in Physics, and M.Sc. and Ph.D. degrees in Electrical Engineering from Université Laval, Québec, P.Q., Canada in 1973, 1975 and 1978 respectively. In 1978, he joined the faculty of the École Polytechnique, Université de Montréal, Montréal, P.Q., Canada, where he is currently a Full Professor. He has been the Head of the Department of Electrical and Computer Engineering from 1996 to 1998 and the Chief Executive Officer of École Polytechnique from 1998 to 2002. He is now the Head of Laboratoire Scribens at this institution. Over the last thirty years, Professor Plamondon has been involved in many pattern recognition projects, particularly in the field of on-line and off-line handwriting analysis and processing. He has proposed many original solutions, based on exhaustive studies of human movement generation and perception, to problems related to the design of automatic systems for signature verification and handwriting recognition, as well as interactive electronic penpads to help children learning handwriting and powerful methods for analyzing and interpreting neuromuscular signals. His main contribution has been the development of a kinematic theory of rapid human movements which can take into account, with the help of a unique basic equation called a delta-lognormal function, the major psychophysical phenomena reported in studies dealing with rapid movements. The theory has been found successful in describing the basic kinematic properties of velocity profiles as observed in finger, hand, arm, head and eye movements. Professor Plamondon has studied and analyzed these biosignals extensively in order to develop creative and powerful methods and systems in various domains of engineering. Full member of the Canadian Association of Physicists, the Ordre des Ingénieurs du Québec, the Union national des écrivains du Québec, Dr Plamondon is an also active member of several international societies. He is a Fellow of the Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS; 1989), of the International Association for Pattern Recognition (IAPR; 1994) and of the Institute of Electrical and Electronics Engineers (IEEE; 2000). From 1990 to 1997, he was the President of the Canadian Image Processing and Pattern Recognition Society and the Canadian representative on the board of Governors of IAPR. He has been the President of the International Graphonomics Society (IGS) from 1995 to 2007. He has been involved in the planning and organization of numerous international conferences and workshops and has worked with scientists from many countries. He is the author or co-author of more than 300 publications and owner of four patents. He has edited or co-edited four books and several Special Issues of scientific journals. He has also published a children book, a novel and three collections of poems.

Julian Fierrez received the M.Sc. and the Ph.D. degrees in Electrical Engineering from Universidad Politecnica de Madrid, Madrid, Spain, in 2001 and 2006 respectively. Since 2002 he has been affiliated with the Biometric Recognition Group – ATVS, first at Universidad Politecnica de Madrid, and since 2004 at Universidad Autonoma de Madrid, where he is currently an Associate Professor. From 2007 to 2009 he was a visiting researcher at Michigan State University in USA under a Marie Curie fellowship. In the past, he has also conducted 3-month research visits at Halmstad University in Sweden (2003), Bologna University in Italy (2004), Michigan State University in USA (2005), and University of Surrey in UK (2006). His research interests and areas of expertise include signal and image processing, pattern recognition, and biometrics, with emphasis on signature and fingerprint verification, multibiometrics, biometric databases, and system security. Dr. Fierrez has been and is actively involved in European projects focused on biometrics (e.g., FP6 BioSec IP, FP6 BioSecure NoE, and FP7 BBFor2 Marie Curie ITN), and is the recipient of a number of research distinctions, including: best poster paper at AVBPA 2003, Rosina Ribalta award to the best Spanish Ph.D. proposal in ICT in 2005, best Ph.D. thesis in computer vision and pattern recognition in 2005–2007 by the IAPR Spanish liaison (AERFAI), Motorola best student paper at ICB 2006, EBF European Biometric Industry Award 2006, and IBM best student paper at ICPR 2008.

Javier Ortega-Garcia received the M.Sc. degree in Electrical Engineering (Ingeniero de Telecomunicacion), in 1989; and the Ph.D. degree “cum laude” also in Electrical Engineering (Doctor Ingeniero de Telecomunicacion), in 1996, both from Universidad Politecnica de Madrid, Spain. Dr. Ortega-Garcia is founder and co-director of the Biometric Recognition Group — ATVS. He is currently a Full Professor at the Escuela Politecnica Superior, Universidad Autonoma de Madrid, where he teaches Digital Signal Processing and Biometric Recognition courses. His research interests are focused on biometrics signal processing: speaker recognition, fingerprint recognition, on-line signature verification, data fusion and multibiometrics. He has published over 150 international contributions, including book chapters, refereed journal and conference papers. He chaired “Odyssey-04, The Speaker Recognition Workshop” (co-sponsored by IEEE), and co-chaired “ICB-09, the 3rd IAPR International Conference on Biometrics”. He has been appointed as Chair of “ICB-13, the 5th IAPR International Conference on Biometrics”.