Improving the Enrollment in Dynamic Signature Verification with Synthetic Samples

Javier Galbally, Julian Fierrez, Marcos Martinez-Diaz, and Javier Ortega-Garcia
Biometric Recognition Group–ATVS, EPS, Universidad Autonoma de Madrid
{javier.galbally, julian.fierrez, marcos.matinez, javier.ortega}@uam.es

Abstract

A novel scheme to generate multiple synthetic samples from a real on-line handwritten signature is proposed. The algorithm models a transmission channel which introduces a certain distortion into the real signature to produce the different synthetic samples. The method is used to increase the amount of data of the clients enrolling on a state-of-the-art HMM-based signature verification system. The enhanced enrollment results in performance improve up to 70% between the case in which only one real sample of the user was available for the training, and the case where the proposed algorithm was used to generate additional synthetic training data.

1. Introduction

With the increasing importance that biometric security systems are acquiring in today's society and their introduction in many daily applications, a growing interest is arising for the generation of synthetic biometric traits such as voice, fingerprints [2], iris [14], handwriting [9], or signature [8]. The generation of these synthetic samples is of great interest as can help to overcome the lack of biometric data existing in many applications specially at the enrollment stage.

It should be emphasized that, although there are multiple works which address the problem of generating synthetic traits, 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 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. Therefore, this class of methods can be helpful to synthetically augment the size of the enrollment set of signatures in identification and verification systems, a critical parameter in signature biometrics [6].

  The great majority of existing approaches for synthetic signature generation are based on this type of strategy [11, 3, 13]. This is the approach that we will follow in the present work.

- **Combination of different real samples**. This is the approach followed by most speech and handwriting synthesizers. This type of algorithms start from a pool of *real* n-grams (e.g., isolated letters, or combination of two or more letters) and using some type of concatenation procedure combine them to form the synthetic samples [1, 9]. 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 totally new synthetic individuals.

- **Synthetic-individuals**. In this case, some kind of *a priori* knowledge about a certain biometric trait (e.g., minutiae distribution, iris structure, signature length, etc.) is used to create a model that characterizes that biometric trait for a population of subjects. New *synthetic* individuals can then be generated sampling the constructed model. Different model-based algorithms have been presented in the literature to generate synthetic individuals for biometric traits such as iris [14], fingerprint [2], and signature [8, 4].

In the present article we describe a novel approach for generating duplicated samples based on modeling the variability between the signatures of a given user. This estimated user intravariability is then applied as a distortion that is introduced to the input real signature, in order to obtain the duplicated samples. The algorithm is in this way used to increase the amount of enrollment data. Experiments are conducted on a state-of-the-art HMM-based signature recognition system, showing that using synthetically

978-0-7695-3725-2/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDAR.2009.38

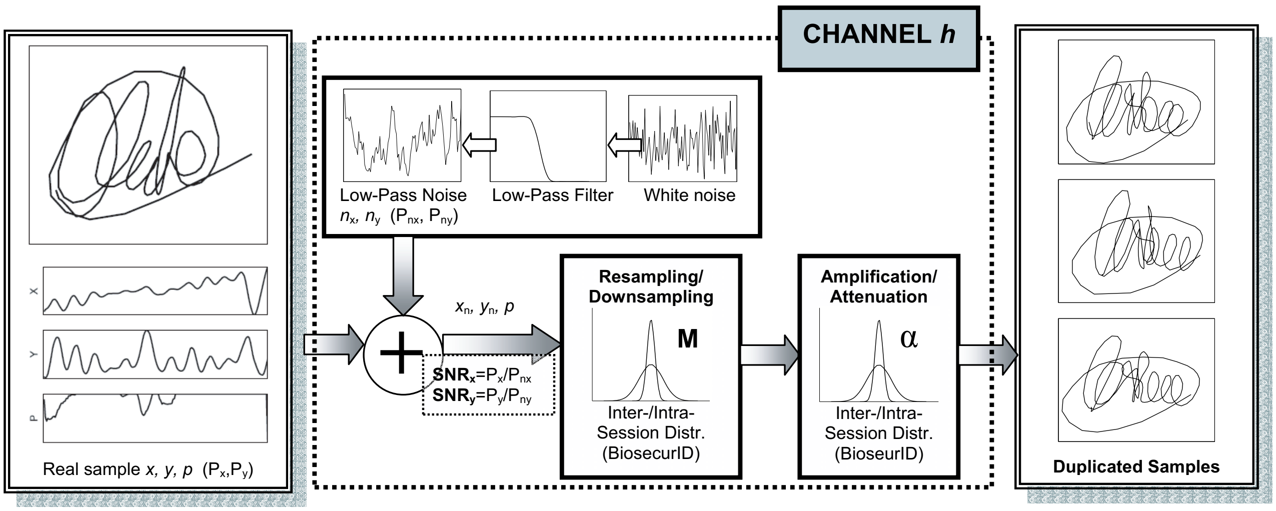

Figure 1. General architecture of the algorithm for generating duplicated samples.

generated signatures for enrollment as a complement to the real data drastically improves the system performance (up to 70% improvement in some of the scenarios considered).

The paper is structured as follows. The algorithm for generating synthetic signatures is presented in Sect. 2. The development and test sets, and the protocol followed in the experiments, are described in Sect. 3. In Sect. 4 we show and analyze the results obtained. Conclusions are finally drawn in Sect. 5.

2. Generating duplicated samples

Lets consider the signing process as follows. A clean dynamic signature $[x(t), y(t), p(t)]$, 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. Under this framework, the generation of multiple samples from a given clean signature is straightforward given by the distortion parameters.

In the present work we consider three different stages to model the distortions introduced by the channel h in the signature time signals: i) noise addition according to a particular Signal to Noise Ratio (SNR), ii) resampling/downsampling of the original signal by a factor M, and iii) amplification/attenuation of the signal in terms of a parameter α. Next we describe each of the three distortion stages (see Fig. 1).

- **Noise addition** (SNR). Low-pass 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., $\text{SNR}_x = P_x/P_{nx}$). 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.

  In this step of the algorithm 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 intrasession ($M \in [-M^{\text{intra}}, M^{\text{intra}}]$) or intersession ($M \in [-M^{\text{inter}}, M^{\text{inter}}]$) variability, being in general $|M^{\text{intra}}| < |M^{\text{inter}}|$.

- **Amplification/Attenuation** (α). An affine scaling is finally applied to all three signals according to a parameter α (which varies for each time function) [10]. Analogously to the resampling parameter M, the amplification factor α follows a uniform distribution between $[-\alpha_x^{\text{intra}}, -\alpha_x^{\text{intra}}]$ for intrasession samples, and between $[-\alpha_x^{\text{inter}}, -\alpha_x^{\text{inter}}]$ for intersession samples (similarly for functions y and p). For a given value of the parameter α_x, the scaled function x_α is computed as $x_\alpha = (1 + \alpha_x)x$.

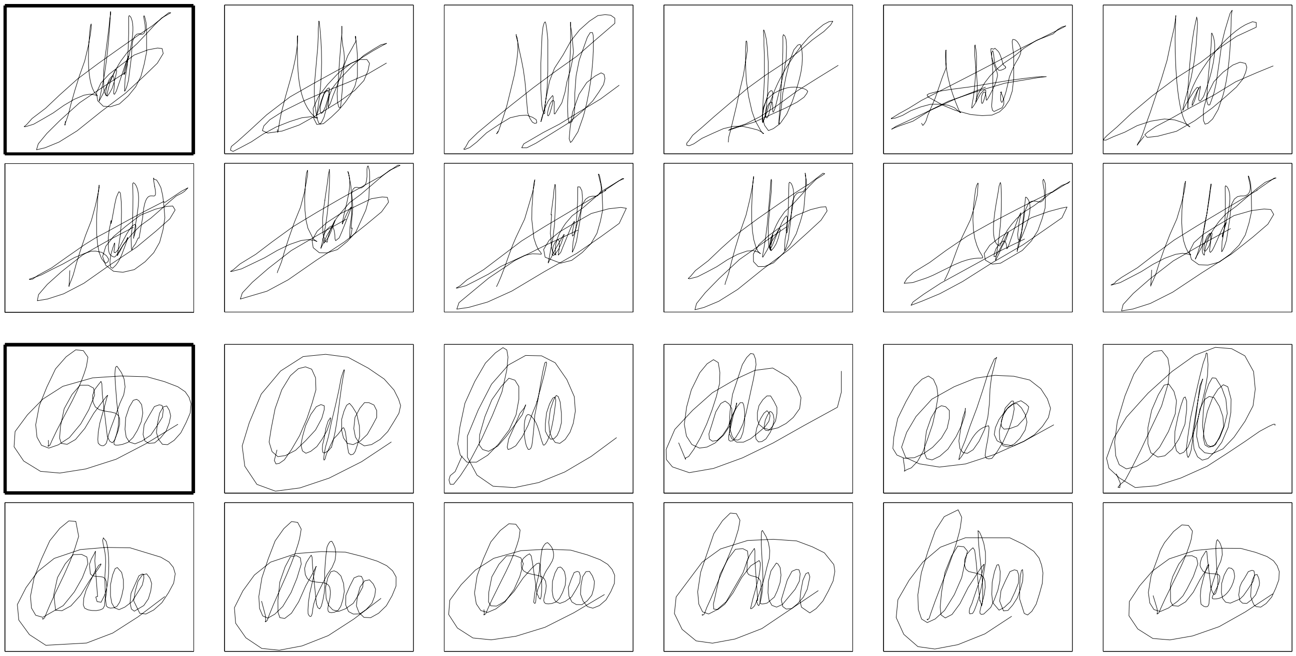

Figure 2. Real (odd rows) and synthetic (even rows) samples of two different users of MCYT. The duplicated samples were generated from the real signature highlighted with a thicker frame.

3. Experimental protocol

In order to avoid biased results, two totally different datasets were used as development (to estimate the generation model parameters) and test sets (where results on the efficiency of the algorithm are obtained).

For the estimation of the algorithm parameters (SNR, M, and α) we used part of the signature data in the BiosecurID multimodal database [5]. BiosecurID comprises eight different biometric traits of 400 users and was captured in four acquisition sessions over a six month time span (which makes it a very efficient tool to estimate the inter and intrasession variability). The signature subset comprises for each user, 16 original samples (four samples per session), and 12 forgeries carried out with an increasing degree of skill over the sessions (both the off-line and on-line information of each signature is available). In the present work, the imitations were discarded and only the $400 \times 16 = 6,400$ genuine dynamic signatures were used as development set. The values obtained on this dataset for each of the parameters defining our generation model were:

- Parameter SNR. Based on the assumption of uncorrelated signature signals and noise, we estimate the SNR averaging the noise (computed between pairs of genuine signatures avoiding repetitions) across users. Thus, the global SNR of signal x of a specific user ($\mathrm{SNR}_x^{\mathrm{U}}$) is estimated as:

$$\mathrm{SNR}_x^{\mathrm{U}} = \frac{1}{\mathrm{C}(N_{gs},2)} \sum_{k=1}^{N_{gs}} \frac{P_x^i}{|P_x^i - P_x^j|} \quad \text{for} \quad j > i,$$

where N_{gs} represents the number of considered genuine signatures from the user, and $\mathrm{C}(N_{gs},2)$ is the number of possible combinations of the N_{gs} signatures taken in pairs: $\mathrm{C}(N_{gs},2) = N_{gs}!/2!(N_{gs}-2)!$.

The final SNR_x distribution is estimated using the 400 $\mathrm{SNR}_x^{\mathrm{U}}$ measures obtained from BiosecurID.

Parameter SNR_y is computed similarly, being in both cases the genuine pairs of signatures (N_{gs}) either from the same or different acquisition sessions (intrasession and inter-session SNR models, respectively).

The results show that the power of the noise added in the x coordinate to produce intersession samples P_{nx}^{inter} has to be around 8% higher than in the case of intrasession repetitions P_{ny}^{intra} (i.e., $P_{nx}^{\mathrm{inter}} = 1.08 P_{nx}^{\mathrm{intra}}$). In the case of the noise affecting the y coordinate function, the variability between samples captured in the same and different sessions is slightly higher: $P_{ny}^{\mathrm{inter}} = 1.11 P_{ny}^{\mathrm{intra}}$.

Table 1. EER for the HMM-based signature verification system, with and without considering the pressure information, for the random and skilled forgeries scenarios and for different cases of enrollment data. R stands for *Real*, and S for *Synthetic*.

	Without pressure information. EER (%)					
	1R	5R	20R	1R + 4S	1R + 19S	4R + 16S
Random	23.85	5.71	1.81	7.87	7.11	**2.12**
Skilled	32.15	14.57	9.13	16.24	15.60	**10.25**

	With pressure information. EER (%)					
	1R	5R	20R	1R + 4S	1R + 19S	4R + 16S
Random	22.84	4.27	0.87	7.40	6.60	**1.17**
Skilled	31.03	10.97	5.57	16.07	15.60	**6.35**

- Parameter M. The value of the intrasession duration variability found in the development set is defined by $M^{intra} = 0.1$, while the intersession variability follows a uniform distribution characterized by $M^{\text{inter}} = 0.14$.

- Parameter α. The values that define the uniform distributions from which this parameter is extracted are (for the three time functions x, y, and p):
$[\alpha_x^{\text{intra}}, \alpha_x^{\text{inter}}] = [0.06, 0.08]$,
$[\alpha_y^{\text{intra}}, \alpha_y^{\text{inter}}] = [0.08, 0.11]$,
$[\alpha_p^{\text{intra}}, \alpha_p^{\text{inter}}] = [0.05, 0.06]$.

As test set, the dynamic signature data of the MCYT database (comprising signature and fingerprint information of 330 users) was used [12]. The signature dataset is formed by 25 original samples and 25 skilled forgeries per user (captured in five different acquisition sets). These data are used to estimate the performance of an HMM-based signature recognition system (using 12 states and 4 mixtures per state) [7] for both random and skilled forgeries under different conditions of enrollment: using only real samples from the user, or complementing these data with synthetically generated signatures (the different enrollment scenarios will be specified in Sect. 4).

In both cases (skilled and random forgeries) the genuine scores are computed matching the enrollment data with the last 5 original signatures of the user (resulting in $330 \times 5 = 1,650$ similarity scores). The way to obtain the impostor scores differs between both scenarios: i) in the random forgeries case each user's model is compared with one signature of the remaining users (i.e., $330 \times 329 = 108,529$ impostor scores), and ii) when considering skilled forgeries the enrollment data of each user is matched with the 25 imitations of that same user (i.e., $330 \times 25 = 8,250$ impostor scores).

In the odd rows of Fig. 2 we show six real samples of two different users in MCYT. The first two signatures correspond to the first acquisition set, while each of the remaining samples belong to each of the other four sets. In the even rows we depict six synthetic samples corresponding to the same users, generated with the method proposed in this work from the real samples highlighted with a thicker frame. The first two synthetic samples were produced applying intrasession variability, and the other four with intersession parameters.

4. Results

The aim of the experiments is to find if adding synthetically generated samples (according to the model described in the present work) to the real enrollment data of the clients, can improve the performance of signature recognition systems.

For this purpose we evaluated an state-of-the-art HMM-based system under different scenarios for enrollment:

- Using only real samples to compute the enrollment model of each user.

- Complementing the real data of the user with synthetically generated samples.

In particular, we considered the cases of enrolling with 1, 5, and 20 real signatures, and enrolling with 1R+4S (1 Real, 4 Synthetic generated from that real signature), 1R+19S, and 4R+16S (4 synthetic samples produced from each of the 4 real samples). The experiments were carried out with and without taking into account the pressure information, and for both random and skilled forgeries. Results are given in Table 1 in the for of Equal Error Rates (EER in %).

From the results obtained with no pressure information, we can see that in the case of having just one real signature for the enrollment of the client, we can improve the system performance in nearly 70% (from EER=23.84% to 7.87%) for the random forgeries scenario, and nearly 50% (from 32.15% to 16.24%) for the skilled forgeries case, by adding just four synthetic samples generated from that real signature following the proposed method. Furthermore, the EER obtained using five real enrollment signatures from the same session (5.71% in random forgeries, and 14.57% for skilled) is comparable to that obtained using only one real sample complemented with four synthetic samples (7.87% and 16.24% for random and skilled forgeries, respectively).

We can also observe in Table 1 comparing the results 1R+4S with 1R+19S that the EER gain introduced with an increasing number of synthetic samples generated from the same real signature saturates: EER of 7.87% with 1R+4S, to 7.11% with 1R+19S. This fact suggests that the variability modeled by the proposed approach, although very realistic as has been proven in the comparison between 1R, 5R and 1R+4S, is not enough to totally capture the natural signature variability (this is specially evident if we compare 20R with 1R+19S).

To avoid this EER gain saturation we tested the HMM-based recognition system in a 4R+16S enrollment data scenario, where four synthetic samples are generated from each of the four real samples (all taken from the first session as in the 5R case). The results are highlighted in **bold** in Table 1. There we can observe that, even though we are just considering four real signatures, the introduction of additional synthetic samples for training drastically improves the system's EER compared to training with five real samples (over 60% improvement for random forgeries and nearly 30% for skilled forgeries). The results are in this case (4R+16S) totally comparable to the (unrealistic) scenario where the enrolling data comprises 20 real samples (1.81% and 9.13% EER in 20R for random and skilled forgeries, against 2.12% and 10.25% EER for the same cases with 4R+16S).

Although the analysis of the results has been made for the case in which the pressure function was not considered, very similar conclusions can be drawn from the table where this information is taken into account.

5. Conclusions

A novel algorithm to generate multiple samples from real on-line signatures has been proposed. The method models the existing variability between different signature samples of a given user in an analogue way to the distortion introduced by a communication channel to a signal being transmitted. The algorithm has been used to synthetically increase the amount of available enrollment data (complementing the real enrollment signatures of a user with synthetic samples) and evaluated (using totally different development and test sets) on a state-of-the-art HMM-based signature recognition system. The results show that the use of synthetically generated signatures (following the proposed algorithm) drastically improves the system performance with gains of up to 70% in the EER for realistic testing scenarios.

As a result, it is patent that this is a very powerful tool to enhance the performance of automatic signature recognition systems when very few enrollment data is available. Furthermore, it could also be used in different signature and handwriting applications such as: individuality studies, new recognition algorithms, variability compensation methods or template update.

6. Acknowledgements

J. G. is supported by a FPU Fellowship from Spanish MEC, and J. F. is supported by a Marie Curie Fellowship from the EC. This work was supported by Spanish MEC under project TEC2006-13141-C03-03.

References

[1] L. Ballard, D. Lopresti, and F. Monrose. Forgery quality and its implications for behavioral biometric security. *Trans. on Systems, Man, and Cybernetics*, 37:1107–1118, 2007.

[2] R. Cappelli. *Handbook of Fingerprint Recognition*, chapter Synthetic Fingerprint Generation, pages 203–231. Springer, 2003.

[3] M. Djioua, C. O'Reilly, et al. An interactive trajectory synthesizer to study outlier patterns in handwriting recognition and signature verification. In *Proc. ICPR*, 2006.

[4] M. Djioua and R. Plamondon. A new algorithm and system for the characterization of handwriting strokes with delta-lognormal parameters. *Trans. on Pattern Analysis and Machine Intelligence*, 2009. to appear.

[5] J. Fierrez, J. Galbally, et al. Biosecurid: a multimodal biometric database. *Pattern Analysis and Applications*, 2009. To appear.

[6] J. Fierrez and J. Ortega-Garcia. *Handbook of biometrics*, chapter On-line signature verification, pages 189–209. Springer, 2008.

[7] J. Fierrez, J. Ortega-Garcia, et al. HMM-based on-line signature verification: feature extraction and signature modeling. *Pattern Recognition Letters*, 8:2325–2334, 2007.

[8] J. Galbally, J. Fierrez, et al. Synthetic generation of handwritten signatures based on spectral analysis. In *Proc. SPIE BTHI*, 2009. to appear.

[9] A. Lin and L. Wang. Style-preserving english handwriting synthesis. *Pattern Recognition*, 40:2097–2109, 2007.

[10] M. E. Munich and P. Perona. Visual identification by signature tracking. *Trans. PAMI*, 25:200–217, 2003.

[11] C. Oliveira, C. A. Kaestner, et al. Generation of signatures by deformations. In *Proc. BSDIA*, pages 283–298, 1997.

[12] J. Ortega-Garcia, J. Fierrez-Aguilar, et al. MCYT baseline corpus: a bimodal biometric database. *IEE Proc. VISP*, 150(6):391–401, 2003.

[13] C. Rabasse, R. M. Guest, and M. C. Fairhurst. A method for the synthesis of dynamic biometric signature data. In *Proc. ICDAR*, 2007.

[14] J. Zuo, N. A. Schmid, et al. On generation and analysis of synthetic iris images. *IEEE Trans. IFS*, 2:77–90, 2007.