Reducing the Template Aging Effect in On-Line Signature Biometrics

Ruben Tolosana, Ruben Vera-Rodriguez, Julian Fierrez, Javier Ortega-Garcia

[1] *BiDA Lab - Biometrics and Data Pattern Analytics Laboratory, Universidad Autonoma de Madrid, 28049 Madrid, Spain*

** E-mail: ruben.tolosana@uam.es*

Abstract: On-line signature recognition is an area of growing interest in recent years due to the massive deployment of high-quality digitizing tablets, smartphones, and tablets in many commercial sectors such as banking. In addition, handwritten signature is one of the most socially accepted biometric traits as it has been used in financial and legal agreements for over a century. In this current environment for signature biometrics, the number of stored samples or templates per user can grow very fast, making it possible to train more robust statistical user models, improving the performance of the biometric systems and in particular reducing the template aging effect. This paper carries out an exhaustive experimental analysis of template update strategies for three well known on-line signature verification approaches, extracts various practical findings related to the template aging effect in signature biometrics, and configures time-adaptive improved versions of the considered baseline approaches overcoming to some extent the template aging. Our improved approach achieves system performances of 2.1% and 0.2% Equal Error Rate for skilled and random forgery cases, respectively. These results show the efficacy of our methodology.

1 Introduction

Due to the technological evolution and the significant improvement of sensors' quality, devices such as smartphones and tablets have experimented a great deployment in the last recent years [1]. The use of these new devices has enabled the deployment of automatic biometric authentication systems in many different sectors [2] due to the many advantages compared to traditional ways of authentication (i.e., based on passwords or cards). The handwritten signature is one of the most socially accepted traits as it has been used in financial and legal agreements scenarios for many years [3, 4]. In addition, it is worth noting that signatures are very easy to acquire by means of these devices through stylus or even using the finger as the writing tool [5-8].

One of the main challenges in signature verification is related to signature variability. While genuine signatures can differ significantly (high intra-class variability), skilled forgeries could be similar to genuine signatures (low inter-class variability) [9-11]. Another important problem related to the intrinsic or intra-class variability of signatures is known as *biometric aging* [12]: the gradual degradation of a system performance due to the changes suffered by the user's trait along the time. Finally, it is also worth considering the sources of extrinsic variability present in device interoperability scenarios [13] due to the high deployment of many different devices in the last years.

The motivation behind the work presented here is to develop an on-line signature verification system suitable for banking applications. User approval of different banking transactions has been performed for many decades using the handwritten signature, traditionally on paper (off-line signature). For the last years, banks around the world have started to digitize all documentation in order to 1) reduce the use of paper, and 2) have all the documentation easily accessible on-line. Recently, banks are introducing digitizing tablets for the users to perform their signatures in order to make the whole process digital, without the need for scanning the signed documentation on a later stage. At this point, on-line handwritten signatures can be used to perform user authentication as an added layer of security. The usage of on-line signature verification systems in banking is a natural choice since users have used their signatures in banking traditionally. In this work, a thorough analysis of the challenges existing within the on-line signature trait is carried out in order to achieve high accurate, reliable, feasible, and long-term on-line handwritten signature authentication systems for commercial sectors such as, but not only, banking. In fact, we believe that having the challenging operational requirements of banking applications in mind will help to move forward this technology in other fields in which signature biometrics may also find good application.

In the case of banking scenarios, the problem of signature aging is present since there can be a significant time lapse between the signatures acquired in the enrolment session and the signature used to verify the user identity. However, it is common that many samples or templates per user are stored in different sessions making it possible to train biometric systems with a larger amount of information and therefore enabling significant improvements of biometric performance. This scenario has not been analysed properly in previous works mainly due to the small number of samples and sessions of the databases commonly used for research. In this work we analyse this scenario of increasing amounts of signatures making use of an extended version of the ATVS On-Line Signature Long-Term database [12]. The main goal of this work is to analyse template update strategies over popular signature recognition approaches regarding the number of training signatures available to build a user's template and the elapsed time between the training and testing signatures, with the final goal of improving the system performance and reducing the template aging effect. In order to perform a complete template update analysis, the results from our previous work [14] based on setting different optimal system configuration parameters are taken into account in this work.

Some of the most well known approaches for on-line signature biometrics are considered in this work: Hidden Markov Models (HMM), Gaussian Mixture Model (GMM), and Dynamic Time Warping (DTW). Both HMM and DTW have been shown to be quite competitive approaches when combined with proper features [15] and adequate configuration [16-20]. The GMM approach, which can be seen as a particular case of HMM with only one hidden state, is also considered in this work as it has provided good performances in previous works for on-line signature verification [21, 22].

Fig. 1 illustrates the concept of template update as studied here. The Traditional Approach on top only uses for enrolment an initial collection of genuine signatures. In the Present Study (Fig. 1 bottom), we explore ways to incorporate additional enrolment data coming across time.

Based on the main scenario in our mind mentioned before (in-branch banking operations), we assume that signatures coming

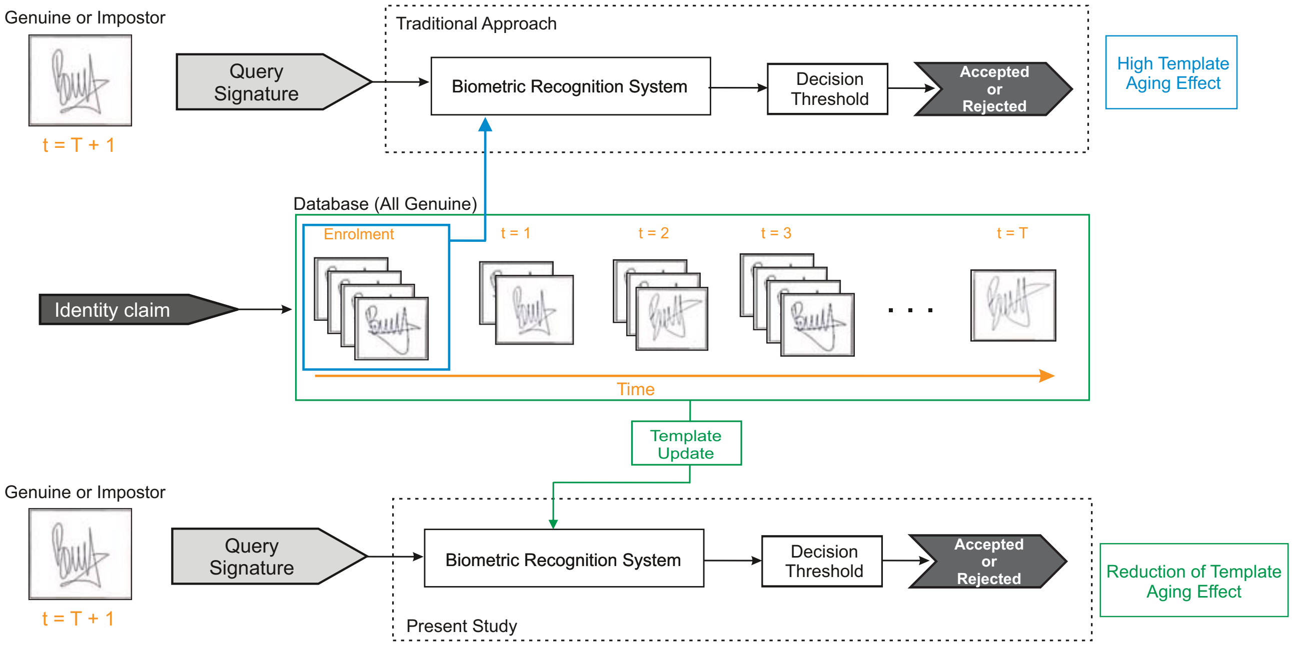


Fig. 1: Template update concept compared to the traditional one based only in an initial collection of enrolment signatures.

across time are all genuine (see Fig. 1) after some kind of human validation (typical when collecting signatures in bank branches, e.g., checking ID cards or knowing the customer). The problem is then authenticating a new signature ($t = T + 1$ in Fig. 1) based on a collection of past genuine signatures. Therefore, the term template update used in this paper should be seen as a supervised template update, since templates are not updated with impostor data [23].

The main contributions of this work are the following:

- This paper reports the first significant experimental results regarding the effect of template aging and template update strategies for on-line signature authentication considering both random and skilled forgeries of the signatures. For this, we have created an extension of the ATVS On-Line Signature Long-Term database in which skilled forgeries are included [14]. The complete signature database is publicly available at `https://github.com/BiDAlab/xLongSignDB`.
- This work carries out an analysis of different template update strategies for the three state-of-the-art systems considered (DTW, HMM and GMM) taking into account different number of signatures in different sessions, reporting various findings and practical recommendations for successful application of template update in signature biometrics.
- A statistical analysis based on the matching score distributions has been carried out to interpret the experimental results.
- A final fusion of the three studied systems (HMM, GMM and DTW) is carried out reducing significantly the template aging effect, especially for the skilled forgery case.

The remainder of the paper is organised as follows. Sec. 2 summarises related works in template update. Sec. 3 describes the methods studied in this work in order to reduce the template aging effect. Sec. 4 describes the extended version of the ATVS On-Line Signature Long-Term database considered in the experimental work. Sec. 5 describes the three signature systems and the experimental work carried out. Finally, Sec. 6 draws the final conclusions and future work.

2 Related Work

The methods studied in this work for reducing the effect of template aging in on-line signature biometrics are based on template update strategies. Template selection and update strategies have been studied in previous works for other biometric traits [24, 25]. In [24] two different methods were proposed for fingerprint verification in order to select the best k templates out of K available. For both methods considered in [24], the number k was fixed to 5 due to application requirements. In [25], the authors proposed a unified taxonomy in which the various updating strategies can be classified under a common framework based on the learning methodology adopted (i.e., supervised and semi-supervised). In addition, an exhaustive study was conducted for selection and template update strategies for face and fingerprint traits showing the pros and cons of different approaches.

However, very few works have studied the problem of template update for on-line signature verification. In [12] template update was studied considering only the case of random forgeries. In [26], authors studied the impact of signature aging and the effectiveness of using a cross-session training strategy. For that purpose, they acquired a database composed of 6 different sessions. Users were asked to perform their signatures using the finger as the writing tool on their own devices. Regarding the experimental work, they considered a feature-based system whose features were extracted from histograms related to X and Y coordinates, speed, angles, pressure, and their derivatives. Results obtained in that work showed the degradation of the system performance when training and query samples belonged to different sessions. Additionally, they analysed the system performance when signatures from multiple sessions were considered for training, achieving better results compared to the case of using just one session for training. However, in that work the database considered was acquired with a very small time gap between the first and last sessions (i.e., only seven days) being difficult to extrapolate these results to real long-term scenarios (e.g., time gap of several months between the training and query signatures). It would also be difficult to know whether the improvement achieved in that work was produced due to the increasing number of signatures used in the different experiments or due to the reduction of the aging effect. It is also worth mentioning that only the case of random forgeries was considered in that work, as skilled forgeries were not performed in the database. Therefore, we consider necessary to carry out a more exhaustive analysis of template update strategies on real long-term scenarios in order to reduce the aging effect for on-line signature verification.

In this work we consider a scenario where the number of signatures acquired per user can increase rapidly as in real banking or

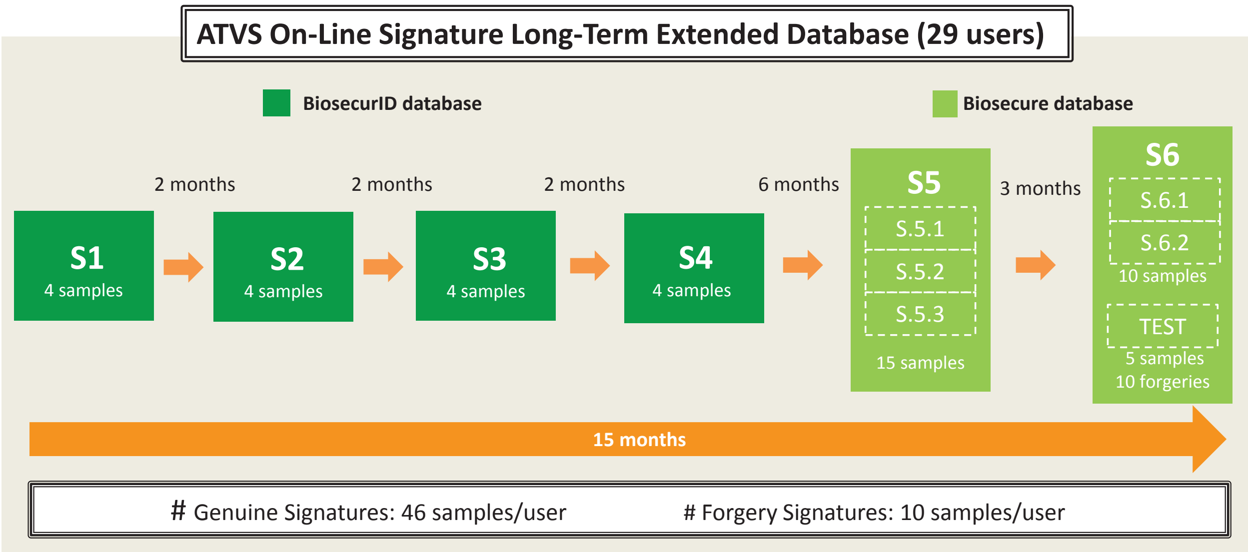


Fig. 2: General time diagram of the different acquisition sessions and number of genuine signatures per user that form the ATVS On-Line Signature Long-Term Extended Database.

commercial applications nowadays. Therefore, these signatures can be used to update users templates and reduce the aging effect. In order to perform a complete template update analysis, the results from our previous work [14] based on setting different optimal system configuration parameters are taken into account in this work as a first stage.

3 Methods

3.1 Template Update Strategies

Two different approaches are analysed in the experimental work for template update:

1. Adding all the available genuine signatures of the subject at hand across time to the ones from the enrolment session.
2. Not considering old genuine signatures from the subject at hand for updating the user model.

Several intermediate configurations are also analysed in order to study the performance evolution of the three systems considered. Besides, it is important to highlight that many additional factors are considered in this stage (i.e., computational cost, resources, etc.) as they are very important for practical applications.

3.2 System Complexity Configuration

Finding an optimal system configuration for a given task can provide significant improvements of performance. In [27], a preliminary analysis of the system performance was carried out considering different system configuration parameters for an HMM-based on-line signature verification system. The main limitation of that work was that only 5 training signatures were considered in the user models to study the optimal system configuration parameters, so a broader study including different number of training signatures per user model will be helpful.

In our previous work [14], an exhaustive analysis of the system performance was carried out considering different system configuration parameters for HMM- and GMM-based systems in scenarios where the number of available training signatures per user increases with time. Table 1 summarises the optimal system configuration parameters proposed in [14], which are considered in this work as a starting point in order to perform a more comprehensive analysis of template update strategies for the HMM and GMM systems. In Table 1 we see that when the number of available training signatures

Table 1 Optimal system configuration parameters regarding the number of available training signatures. N denotes the number of hidden states and M the number of Gaussian mixtures per state.

	HMM		GMM
# Training Signatures	N	M	M
<15	2	16	32
16 to 31	32	2	128
>31	64	2	512

is small, the optimal system configuration for the HMM system is based on a small number of hidden states (N = 2) and a medium number of Gaussian mixtures per state (M = 16). On the other hand, as the number of available signatures increases (between 16 and 31 signatures), then the number of optimal hidden states increases (N = 32) and the number of mixtures per state decreases to M = 2. Finally, the number of hidden states increases up to N = 64 for the case of having more than 31 available signatures. For a GMM-based system, as the number of training signatures increases, the number of Gaussian mixtures also increases (M = 512 for 41 training signatures).

3.3 Statistical Analysis

For interpreting our results we have applied a statistical analysis similar to previous works [28]. In that work authors proposed a metric to measure the quality of an on-line signature template derived from a set of enrolled signature samples in terms of its distinctiveness against random forgeries. The use of random and not skilled forgeries for measuring our proposed template quality (Q) is motivated due to the lack of skilled forgeries in real scenarios for training.

Let (μ_g,σ_g) and (μ_r,σ_r) be the mean and standard deviation of the genuine and random matching score distributions provided by the on-line signature verifier, then the template quality for these two distributions is defined as follows:

$$Q = \frac{\|\mu_g - \mu_r\|}{\sqrt{(\sigma_g^2 + \sigma_r^2)/2}} \quad (1)$$

The goal of this template quality metric Q is to measure how separated are the genuine from the random matching score distributions. The larger the separation between the score distributions, the higher is Q (note that Q is equivalent to the metric d defined in [29]). In this work, we compute both the Equal Error Rate (EER) and this Q metric in order to analyse the different template update strategies proposed.

Table 2 Experimental protocol designed to study the template aging effect (Sec. 5.3.1), and template update strategies (Sec. 5.3.2 and 5.3.3). p/s indicates de number of signatures used per session.

Experiment	Training	# Signatures	# Sessions	Aging Analysis	Template Update
A	S1	4	1	X	
B	S2	4	1	X	
C	S3	4	1	X	
D	S4	4	1	X	
E	S.5.2	4	1	X	
F	S.6.2	4	1	X	
G	S1	4	1		X
H	S1-S4	16	4		X
I	S1-S5	31	5		X
J	S1-S4, S.5.2	20 (4 p/s)	5		X
K	S2-S4, S.5.2	16 (4 p/s)	4		X
L	S3, S4, S.5.2	12 (4 p/s)	3		X
M	S4, S.5.2	8 (4 p/s)	2		X
N	S.5.2	4	1		X
O	S5	15	1		X

4 Signature Database

The database used to carry out the experimental work of this paper is an extended version of the ATVS On-Line Signature Long-Term database [12]. Fig. 2 shows the number of genuine signatures per user and the general time diagram of the different acquisition sessions of it. This database was used in [12] taking into account only random forgeries. Skilled forgeries of these signatures have been also included in this extended version of the database, which is publicly available at `https://github.com/BiDAlab/xLongSignDB`. This database comprises a total of 29 users. The inter-session variability problem is also considered in this database as signatures were acquired in 6 different sessions (S1 to S6 in Fig. 2) within a 15-month time span. Sessions from S1 to S4 are composed of 4 genuine signatures per user and have a two month interval between them (i.e., BiosecurID Signature Subset [30]). The acquisition of S5 and S6 sessions was performed in a different campaign (i.e., Biosecure Signature Subset [31]) which started 6 months after the first campaign had finished. It comprises 30 genuine signatures per user distributed in two acquisition sessions separated three months. Therefore, the total number of genuine signatures and skilled forgeries per user are 46 and 10 respectively. To perform skilled forgeries, the users had visual access to the dynamics of the signing process of the signatures they had to forge as many times as they wanted.

Signatures were captured using a pen tablet WACOM Intuos3 A6 digitizer at 100 Hz and writing on a paper. The available information of this device is the following: X and Y pen coordinates, pressure, pen angular orientation (azimuth and altitude angles) and timestamp information. For more information about the ATVS On-Line Signature Long-Term database see [12].

5 Experiments

5.1 On-Line Signature Verification Systems

Three well known systems based on previous studies are considered here: HMM, GMM and DTW. In all of them, signals captured by the digitizer (only X and Y coordinates and pressure) are used to extract a set of 23 time functions for each signature [14]. Time functions related to pen angular orientation (azimuth and altitude angles) were discarded in order to consider the same set of time functions that we would be able to use in general purpose devices such as tablets and smartphones. For the HMM- and GMM-based systems considered in this work, the optimal subset of time-functions used in the experiments is based on [22] whereas for the DTW-based system, the optimal subset of time-functions is based on [13]. Both subsets comprise 9 time functions and were generated via feature subset selection (SFFS) [32].

A final fusion of the three systems after applying template update is also performed computing the sum of the matching scores. The sum rule fusion algorithm has been considered in this work as it is one of the most successful and easiest approaches used in many related works [33, 34]. Before applying the fusion, the scores from each system were normalised to a common range [0,1] using tanh-estimators [35].

5.2 Experimental Protocol

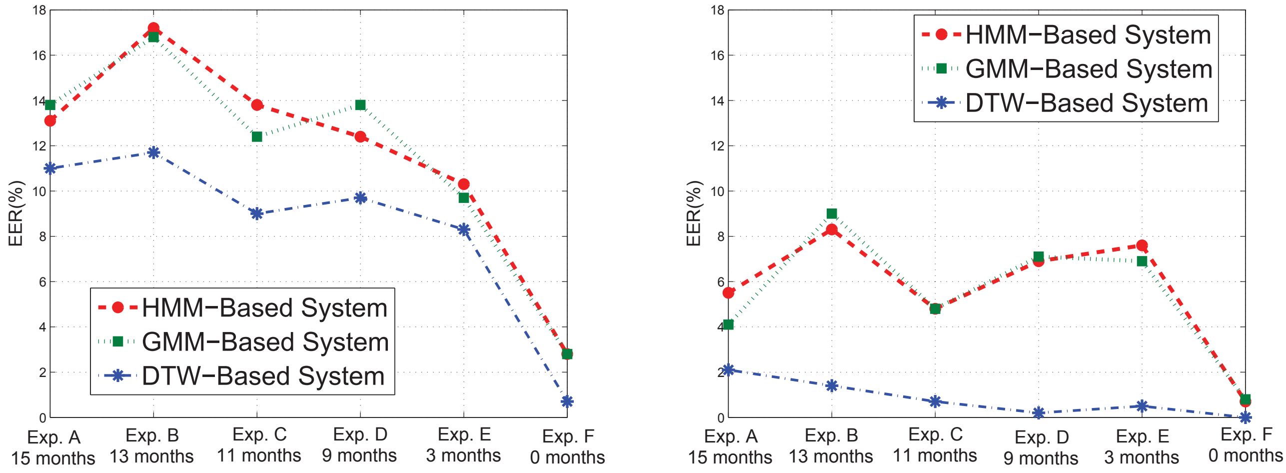
The experimental protocol has been designed to enable the study of the template aging effect and template update strategies for on-line signature authentication. For this, the extended version of the ATVS On-Line Signature Long-Term database is divided into several training sets (in order to analyse different cases and obtain optimal strategies for each one), but only one test set comprised of the last 5 genuine samples and 10 skilled forgeries samples (i.e., TEST block in Fig. 2). This way, fair comparative analysis can be carried out as all experiments use the same signatures for test. Skilled forgery scores are obtained by comparing training signatures against the 10 available skilled forgeries for the same user whereas random or zero-effort forgery scores are obtained by comparing the training signatures to one genuine signature of the remaining users.

First, Sec. 5.3.1 performs an analysis of the template aging effect in on-line signature verification by comparing sets of training data from different sessions with the test set. Second, Sec. 5.3.2 and 5.3.3 carry out an exhaustive search of combinations of training data in order to find an optimal template update strategy for each of the systems considered. All experiments considered in this work for studying the aging effect and template update strategies are depicted in Table 2, which details the number of signatures used for training and the session(s) they come from. A final fusion of the three optimal systems is carried out in Sec. 5.3.4 in order to provide an improved system performance and reduce the template aging effect.

5.3 Experimental Results

5.3.1 ***Template Aging Analysis:*** The aim of this section is to analyse the template aging effect for on-line signature verification. Thus, six different experiments (Exp. A to Exp. F, as depicted in Table 2) have been considered. In all cases four signatures from different sessions are used for training. Exp. A contains training signatures from the first session (S1) with a 15-month time gap with the test session. For the following experiments the time gap (in months) between the training data and the test are 13, 11, 9 and 3 months for Exp. B, C, D, and E, respectively. Finally in Exp. F signatures from the same session (S6) are used for training and test, so the time gap in this case is just a few minutes.

Experiments have been conducted for the three systems considered (i.e., DTW, HMM and GMM). The system configuration



(a) Skilled forgery cases

(b) Random forgery cases

Fig. 3: ***Template Aging Analysis***. Below each experiment we indicate the time gap between training and testing.

parameters of the HMM-based system chosen for this aging analysis are $N = 2$ and $M = 16$, whereas for the GMM-based system we use $M = 32$, being these system configuration parameters the optimal ones depicted in Table 1.

Fig. 3 shows the performance of the systems for all the experiments and for both skilled and random forgeries. Analysing the skilled forgery cases in Fig. 3(a), a general improvement of the performance is achieved for the three systems considered when the elapsed time between the testing and training signatures is reduced. For example, the average performance in terms of EER (%) of the three systems taking into account Exp. A is 12.6% whereas for Exp. E is 9.4%.

However, when analysing random forgery cases (Fig. 3(b)) the effect of the elapsed time does not affect in the same way the performance of the three systems. The performance of the DTW-based system keeps improving as the time between training and testing signatures is reduced (i.e., from 2.1% EER for Exp. A to 0.5% EER for Exp. E). However, for HMM and GMM, the performance does not improve as the time gap is reduced, only showing a very significant improvement of performance for Exp. F, in which the data used for training and testing comes from the same session.

In addition, it is important to highlight that DTW achieves much better performance than HMM and GMM for all the experiments in these conditions. This is due to the fact that DTW is an elastic technique whereas HMM and GMM are statistical algorithms. Therefore, as the number of training signatures considered in these experiments is small (i.e., 4 signatures), it makes sense that DTW works better than HMM and GMM systems, which agrees with previous works [36]. On the other hand, for an increasing volume of enrolment data, as happens in some of the setups explored in Sec. 5.3.2, we will see that the statistical models HMM and GMM are superior to DTW.

Finally, it is also worth noting the results of Exp. F. This experiment does not consider inter-session variability as training and testing signatures come from the same session. In this experiment, the performance for the three considered systems is much better compared to the previous experiments. However, it is important to highlight this is an unusual case as it would only happen in a real application during the enrolment day.

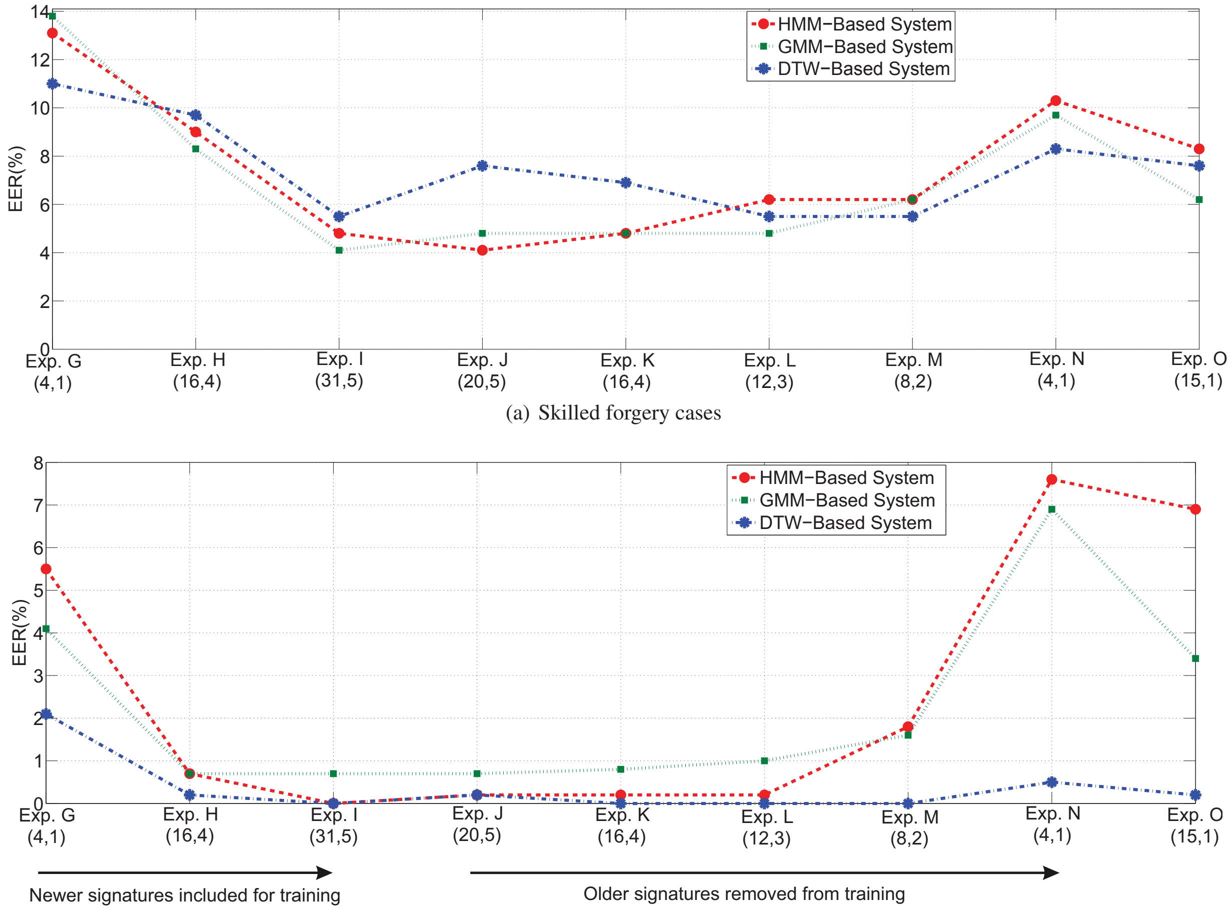
As a general conclusion, we can confirm that on-line signature verification is significantly affected by aging. These trends coincide with previous experiments performed in [12] where signatures from S1 were considered as training signatures and the rest of sessions were used as test, and in [26] where the degradation of the system performance increased when the time lapse between training and test signatures also increased. The goal of the following sections is to reduce the effect of the template aging for on-line signature verification considering different template update strategies regarding the number of available training signatures and the elapsed time between training and testing.

5.3.2 ***Template Update Strategies***: This section focuses on template update strategies given a set of training signatures per user acquired at different sessions, with the final goal of reducing the aging effect. Two different methods are analysed: 1) Adding newer signatures to the enrolment ones; and 2) removing signatures from the older sessions from all the available training signatures. All experiments considered are depicted in Table 2 (Exp. G to Exp. O). The template update strategy followed starts by considering only the enrolment signatures (Exp. G) and adds newer signatures to the enrolled ones (Exp. G to I), this way the time gap between training and test signatures is reduced, and also as the size of the training data increases, better system performance is expected. Then, when all available signatures are considered, we follow the strategy to remove signatures from older sessions (Exp. J to N), in order to analyse whether these signatures with a large time gap with the test ones can still contribute to obtain optimal system performances or not.

It is worth noting that for HMM and GMM, the optimal system configuration parameters described in Sec. 3.2 regarding the number of available training signatures have been taken into account in these experiments in order to perform properly the proposed template update strategies.

Fig. 4 shows the performance of the three systems for all the experiments considered in this section, obtaining this way a global figure to analyse the different possibilities for template update. Results are obtained for both skilled and random forgeries.

Analysing HMM and GMM systems for both skilled and random forgery cases, the performance improves when increasing the size of the templates with newer signatures (from Exp. G to Exp. I). Then, when we remove the older signatures (from Exp. J to Exp. N), the EER increases slowly achieving significantly worse results for Exp. N. Therefore for both HMM and GMM the best performance is achieved for Exp. I with 4.8% and 4.1% EER for HMM and GMM systems respectively for skilled forgery cases and, 0.01% and 0.7% EER respectively for random forgery cases. However, in Exp. I, we are considering 15 signatures from Session S5, which is very unlikely that this happens in a realistic scenario. Therefore, we consider the case of Exp. J, where only four signatures from S5 are used (S.5.2). In this experiment (Exp. J), a few training signatures from different sessions are considered, achieving very similar results compared to Exp. I. Anyway, the process of model training for both HMM and GMM is performed off-line, so the score computation time would not be very affected in the case of having a much larger database with higher number of sessions and signatures.



(a) Skilled forgery cases

(b) Random forgery cases

Fig. 4: ***Template Update Strategies.*** Below each experiment in brackets the first number indicates the number of training signatures, and the second the number of sessions they come from. Exp. G corresponds to using 4 training signatures from the enrolment session. From Exp. G to Exp. I we add for training signatures from more recent sessions. Exp. J has 4 training signatures from each of the 5 sessions. Then from Exp. J to Exp. N we remove signatures from older sessions. Exp. O is included for completeness and contains 15 signatures from the closest session to the test.

Another finding worth highlighting is that it is better to build the user's template considering training signatures coming from different sessions (i.e., Exp. K) instead of using all of them from only one session closer in time (i.e., Exp. O), as there is a significant worsening of performance in this last case. Therefore, the best strategy for both HMM and GMM systems for template update would be to take into account all available training signatures or at least a few training signatures but from several sessions in order to generate a more reliable user's template. This conclusion agrees with the results obtained in [26], in which the on-line signature verification system further improved for an increasing number of training signatures.

On the other hand, the optimal template update strategy for the DTW system is different compared to the HMM and GMM as it can be seen in Fig. 4. Analysing the performance of DTW for both skilled and random forgery cases, the best configurations correspond to Exp. I, L and M with 5.5% EER for skilled forgery cases and 0.01% EER for random forgeries. The first case considers all available training signatures, but the other two just consider 12 and 8 training signatures respectively from closer sessions to the test. In this case the trends would suggest to choose Exp. L and M as the EER increases slowly when we add older training signatures (i.e., Exp. J and K). It is important to highlight that DTW-based systems carry out one to one comparisons of the signatures, so as the number of training signatures increases, the number of DTW comparisons also increases. Thus, it is necessary to establish a limit of comparisons in order to make this system feasible for real time scenarios. As a conclusion for the DTW-based system, the optimal template update strategy would be to consider a few training signatures (i.e., between 8 and 12) from the last sessions closer in time to the test in order to achieve both optimal performance and feasible computation cost.

Finally, in order to quantify the reduction of the aging effect achieved, our proposed template update strategy is evaluated on two different databases: *i)* the ATVS Signature Long-Term Extended database presented in this study, acquired in a 15-month total time span, and *ii)* the remaining 371 users of the BiosecurID database [30], which is composed of four different acquisition sessions (from S1 to S4 in Fig. 2) with a total 6-month time span. It is important to remark that the 371 users of the BiosecurID database have not been used during the analysis of our proposed template update strategy. Table 3 shows the performance of our Proposed Systems incorporating template update strategies for both ATVS Signature Long-Term Extended database and BiosecurID database, respectively. We also include the performance of the traditional case in signature verification (i.e., Baseline) just using the enrolment data from the first acquisition session (S1 in Fig. 2), which would be the case where the aging effect is more pronounced.

Analysing in Table 3 (top) the results obtained for the ATVS Signature Long-Term Extended database, the systems proposed in this work achieve a significant improvement of performance, hence a significant reduction of the template aging effect with an average relative improvement in comparison to the baseline system of 62.0% and 92.2% EER for skilled and random forgeries, respectively.

Results on the unseen users of the BiosecurID database are depicted in Table 3 (bottom). Regarding the experimental protocol, the 4 genuine signatures from the last session (i.e., S4) are always

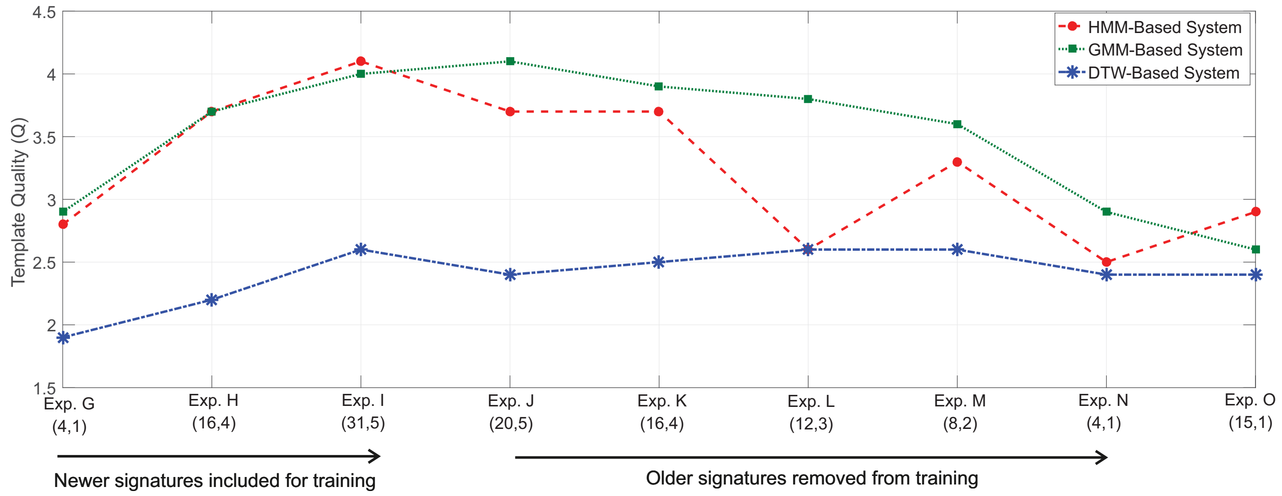


Fig. 5: **Statistical Analysis.** The template quality metric Q is computed for all experiments from Sect. IV.C.2. Below each experiment in brackets the first number indicates the number of training signatures, and the second the number of sessions they come from. Exp. G corresponds to using 4 training signatures from the enrolment session. From Exp. G to Exp. I we add training signatures from more recent sessions. Exp. J has 4 training signatures from each of the 5 sessions. Then from Exp. J to Exp. N we remove signatures from older sessions. Exp. O is included for completeness and contains 15 signatures from the closest session to the test.

Table 3 Comparison of the system performance in terms of EER(%) for Baseline, and Proposed Systems. S stands for Skilled forgeries and R for Random forgeries.

	HMM		GMM		DTW	
	S	R	S	R	S	R
ATVS Signature Long-Term Extended DB						
Baseline	13.1	5.5	13.8	4.1	11.0	2.1
Proposed	4.1	0.2	4.8	0.7	5.5	0.01
BiosecurID DB (+10,000 signatures from 371 users)						
Baseline	10.0	3.8	11.1	4.1	5.8	0.7
Proposed	5.9	2.9	6.4	1.7	3.6	0.2

used as test signatures. For the Baseline system, we just use the 4 genuine signatures from the enrolment session (i.e., S1) for training. However, following our proposed template update strategy, for the HMM and GMM proposed systems we would select all available genuine signatures from S1 to S3 (i.e., 12) for building the user models whereas for the DTW system we would select up to 12 genuine signatures from the last sessions in time, i.e., all available genuine signatures from S1 to S3 as well. Our proposed template update strategy has proved to be very effective against the aging effect, achieving an average relative improvement in comparison to the baseline system of 40.9% and 44.2% EER for skilled and random forgeries, respectively.

5.3.3 ***Statistical Analysis:*** Here we explore the template quality metric Q defined in Sec. 3.3. It is important to highlight that only random forgeries are considered in this statistical analysis due to the lack of skilled forgeries in real scenarios. Fig. 5 shows the template quality Q of the three systems for the same experiments considered in Sec. 5.3.2, obtaining this way a global figure to support the conclusions extracted in Sec. 5.3.2. It is important to remark that the higher the Q value is, the better the template update strategy will be.

Analysing the HMM and GMM systems, the trend of the Q value is to increase with the number of training samples (from Exp. G to Exp. I) and then to decrease when we remove the older signatures (from Exp. J to Exp. N). These results make sense as both HMM and GMM are statistical approaches and they are able to better model the intra-user variability when increasing the number of training signatures and sessions. Therefore, for both HMM and GMM systems the highest value of Q is obtained for Exp. I and J respectively when training signatures from 5 different sessions are considered. This statistical analysis agrees with the template update strategies proposed in Sec. 5.3.2 where the best system performance is obtained when all available training signatures or at least a few training signatures from several sessions are considered.

Analysing the DTW system, the best value of Q is obtained for Exp. I, L and M. In this case the trend would suggest to choose Exp. L or M as the best template update strategies as the DTW system carries out one to one comparisons of the signatures and the larger the number of training signatures is, the higher the computational cost. These statistical results also support the template update strategies proposed in Sec. 5.3.2 for the DTW-based system being the best approach to select a few training signatures (i.e., between 8 and 12) from the last sessions closer in time to the test.

In summary, this statistical analysis based on a template quality metric agrees with the results achieved in the previous sections.

5.3.4 ***Fusion of the Proposed Systems:*** The fusion of biometric systems has been considered in many related works [33, 37] as an easy and reliable way of achieving a further system performance improvement. In this section, the main goal is to carry out the fusion of the three systems studied in order to achieve an improvement of recognition performance and to reduce the template aging effect even further, especially for the challenging case of skilled forgeries. The final fusion is carried out at the score level with the sum rule after normalizing the scores from the three optimal systems to a common range as described in Sec. 5.1. For both HMM and GMM systems, Exp. J has been selected as the optimal template update strategy as it achieves good performance in a realistic set up. In this case, there is a total of 20 training signatures coming from five different sessions. The optimal parameters for the HMM system are N = 32 and M = 2 whereas for the GMM system, M = 128. For the case of the DTW-based system, the optimal template update strategy considered corresponds to Exp. L, which uses a total of 12 training signatures from the last three sessions closer in time to the test. In this case, the final score is the average of all comparisons. The performance of the three systems and the fusion of all of them is represented using DET plots in Fig. 6.

As shown in the figures, the proposed fusion achieves a significant improvement of performance, especially for skilled forgery cases. In this case, the performance of the Fusion System achieves a significant absolute improvement of 2.0% EER compared to the best individual system whereas for the random forgery cases, the proposed fusion does not improve the best system (DTW), which resulted to be almost perfect (i.e., 0.01%). In this case, as the systems being combined behave quite differently, a weighted sum would be more adequate. Anyway, the fused performance is still very competitive (0.2% EER).

6 Conclusions and Future Work

This paper reports the first significant experimental results regarding the effect of template aging and template update strategies for on-line signature authentication considering both random and skilled forgery cases. For this, we have created an extension of the ATVS On-Line Signature Long-Term database, in which skilled forgeries are included. The complete signature database is publicly available at `https://github.com/BiDAlab/xLongSignDB`.

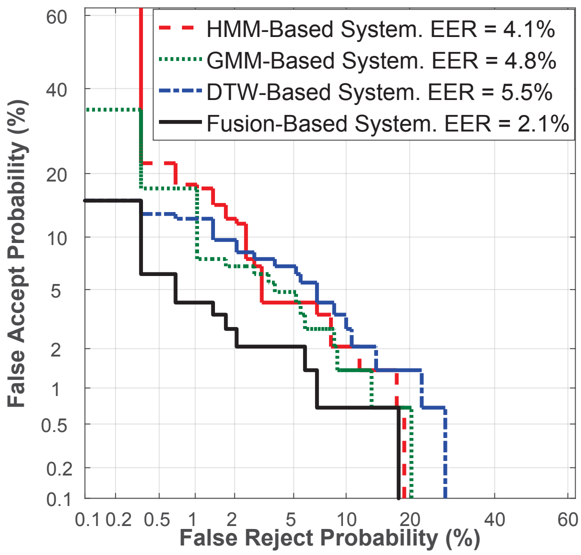
Experiments have been carried out using three well known systems based on time functions: HMM, GMM, and DTW. First, the effect of template aging in on-line signatures has been analysed, concluding that it has a significant impact in the system performance. In order to compensate for this aging effect, an exhaustive experimental analysis of various template update strategies has been carried out considering the results from our previous work [14] as a starting point. For the case of HMM and GMM systems the optimal template update strategy would be to select all available training signatures or at least a few of them from several sessions in order to generate a more reliable user's template. For the DTW system the optimal would be to consider a few training signatures (i.e., between 8 and 12) from sessions closer in time to the test. By incorporating the considered template update techniques, we have demonstrated a significant improvement of performance of the three baseline systems, hence a significant reduction of the template aging effect with similar results to the ideal case for random forgeries, and an average relative improvement of 61.9% EER for skilled forgeries.

Finally, a fusion of the three individual systems after applying the best resulting template update approach has been carried out in order to further improve the recognition performance achieving an EER of 2.1% and 0.2% for skilled and random forgeries respectively. The resulting system described in this work has been deployed successfully in a pilot project in which on-line signature verification will be used massively in the Spanish banking sector.

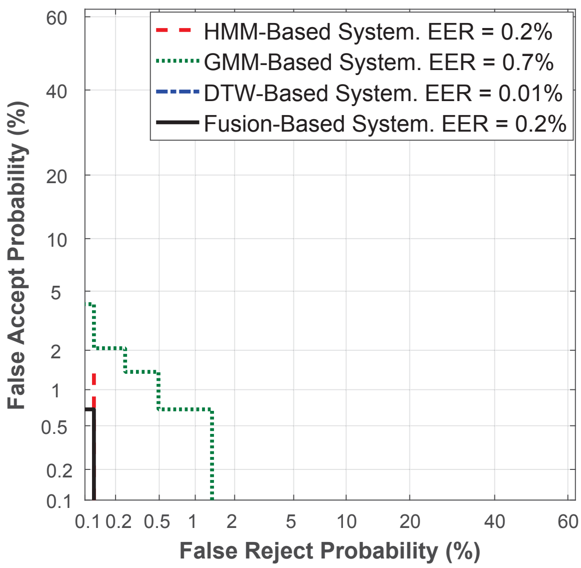
Future work will be oriented to incorporating recent advances in deep learning to the described signature biometrics system [38-40].

7 Acknowledgments

This work has been supported by projects: BIBECA (RTI2018-101248-B-I00 MINECO/FEDER), Bio-Guard (Ayudas Fundación BBVA a Equipos de Investigación Científica 2017) and by UAM-CecaBank. Ruben Tolosana is supported by a FPU Fellowship from Spanish MECD.



(a) Skilled forgery cases



(b) Random forgery cases

Fig. 6: ***Fusion of the Proposed Systems.*** DET curves for the three optimal time functions-based systems after applying the proposed template update approach and fusion of all of them via sum rule of scores.

References

[1] Meng, W., Wong, D., Furnell, S., Zhou, J.: 'Surveying the Development of Biometric User Authentication on Mobile Phones', IEEE Communications Surveys Tutorials, 2015, 17, (3), pp 1268-1293

[2] Mendels, O., Stern, H., Berman, S.: 'User Identification for Home Entertainment Based on Free-Air Hand Motion Signatures', IEEE Transactions on Systems, Man and Cybernetics: Systems, 2014, 44, pp 1461-1473

[3] Impedovo, D., Pirlo, G.: 'Automatic Signature Verification: The State of the Art', IEEE Transactions on Systems, Man, and Cybernetics, Part C, 2008, 38, (5), pp 609-635

[4] Diaz, M., Ferrer, M., Impedovo, D., Malik, M., Pirlo, G., Plamondon, R.: 'A Perspective Analysis of Handwritten Signature Technology', ACM Computing Surveys, 2018, 51, (6), pp 1-39

[5] Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., Ortega-Garcia, J.: 'Benchmarking Desktop and Mobile Handwriting Across COTS Devices: The e-BioSign Biometric Database', PLOS ONE, 2017, 12, (5), p. e69897

[6] Blanco-Gonzalo, R., Sanchez-Reillo, R., Miguel-Hurtado, O., Liu-Jimenez, J.: 'Performance Evaluation of Handwritten Signature Recognition in Mobile Environments', IET Biometrics, 2014, 3, (7), pp. 139-146

[7] Sanchez-Reillo, R.: 'Signature Analysis in the Context of Mobile Devices', Image and Vision Computing, 2016, 55, pp 34-37

[8] Robertson, J., Guest, R.: 'A Feature based Comparison of Pen and Swipe based Signature Characteristics', Human Movement Science, 2015, 43, pp 169-182

[9] Ballard, L., Lopresti, D., Monroe, F.: 'Forgery Quality and Its Implication for Behavioural Biometric Security', IEEE Transactions on Systems, Man and Cybernetics, Part B, 2007, 37, (5), pp 1107-1118

[10] Alonso-Fernandez, F., Fierrez, J., Gilperez, A., Galbally, J., Ortega-Garcia, J.: 'Robustness of Signature Verification Systems to Imitators with Increasing Skills', Proc. International Conference on Document Analysis and Recognition, Barcelona, Spain, 2009

[11] Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Ortega-Garcia, J.: 'Presentation Attacks in Signature Biometrics: Types and Introduction to Attack Detection', Marcel, S., Nixon, M.S., Fierrez, J., Evans, N. (Eds.), 'Handbook of Biometric AntiSpoofing (2nd Edition)', (Springer, 2018)

[12] Galbally, J., Martinez-Diaz, M., Fierrez, J.: 'Aging in Biometrics: An Experimental Analysis on On-Line Signature', PLOS ONE, 2013, 8, (7), p. e69897

[13] Tolosana, R., Vera-Rodriguez, R., Ortega-Garcia, J., Fierrez, J.: 'Preprocessing and Feature Selection for Improved Sensor Interoperability in Online Biometric Signature Verification' IEEE Access, 2015, 3, pp 478-489

[14] Tolosana, R., Vera-Rodriguez, R., Ortega-Garcia, J., Fierrez, J.: 'Update Strategies for HMM-Based Dynamic Signature Biometric Systems', Proc. IEEE International Workshop on Information Forensics and Security, Rome, Italy, 2015

[15] Martinez-Diaz, M., Fierrez, J., Hangai, S.: 'Signature Features', Li, S.Z., Jain, A. (Eds.), 'Encyclopedia of Biometrics' (Springer, 2015), pp 1375-1382

[16] Martinez-Diaz, M., Fierrez, J., Hangai, S.: 'Signature Matching', Li, S.Z., Jain, A. (Eds.), 'Encyclopedia of Biometrics' (Springer, 2015), pp 1382-1387

[17] Liu, Y., Yang, Z., Yang, L.: 'Online Signature Verification Based on DCT and Sparse Representation', IEEE Transactions on Cybernetics, 2015, 45, (11), pp 2498-2511

[18] Song, X., Xia, X., Luan, F.: 'Online Signature Verification Based on Stable Features Extracted Dynamically' IEEE Transactions on Systems, Man and Cybernetics: Systems, 2016, 47, (10), pp 2663-2676

[19] Fischer, A., Plamondon, R.: 'Signature Verification Based on the Kinematic Theory of Rapid Human Movements', IEEE Transactions on Human-Machine Systems, 2016, 47, (2), pp 169-180

[20] Diaz, M., Fischer, A., Ferrer, M.A., Plamondon, R.: 'Dynamic Signature Verification System Based on One Real Signature', IEEE Transactions on Cybernetics, 2016, 48, (1), pp 228-239

[21] Richiardi, J., Drygajlo, A.: 'Gaussian Mixture Models for On-Line Signature Verification', Proc. of the ACM SIGMM Workshop on Biometrics Methods and Applications, Berkeley, California, USA, 2003, pp. 115-122.

[22] Martinez-Diaz, M., Fierrez, J., Krish, R., Galbally, J.: 'Mobile Signature Verification: Feature Robustness and Performance Comparison', IET Biometrics, 2014, 3, (4), pp 267-277

[23] Giot, R., Rosenberger, C., Dorizzi, B.: 'Performance Evaluation of Biometric Template Update', Proc. International Biometric Performance Testing Conference, Gaithersburg, USA, 2012

[24] Uludag, U., Ross, A., Jain, A.: 'Biometric Template Selection and Update: a Case Study in Fingerprints', Pattern Recognition, 2004, 37, (7), pp 1533-1542

[25] Rattani, A.: 'Adaptive Biometric System based on Template Update Procedures', in Dept. of Elect. and Comp. Eng., U. of Cagliari, PhD Thesis, 2010

[26] Sae-Bae, N., Memon, N.: 'Online Signature Verification on Mobile Devices', IEEE Transactions on Information Forensics and Security, 2014, 9, (6), pp 933-947

[27] Fierrez, J., Ortega-Garcia, J., Ramos, D., Gonzalez-Rodriguez, J.: 'HMM-Based On-Line Signature Verification: Feature Extraction and Signature Modeling', Pattern Recognition Letters, 2007, 28, (16), pp 2325-2334

[28] Sae-Bae, N., Memon, N.: 'Quality of Online Signature Templates', Proc. IEEE International Conference on Identity, Security and Behavior Analysis, Hong Kong, China, 2015, pp 1-8

[29] Daugman, J.: 'Biometric Decision Landscapes' University of Cambridge, Computer Laboratory, Tech. Rep. UCAM-CL-TR-482, 2000

[30] Fierrez, J., Galbally, J., Ortega-Garcia, J., *et al.*: 'BiosecurID: A Multimodal Biometric Database', Pattern Analysis and Applications, 2010, 13, (2), pp 235-246

[31] Ortega-Garcia, J., Fierrez, J., Alonso-Fernandez, F., *et al.*: 'The Multi-Scenario Multi-Environment BioSecure Multimodal Database (BMDB)', IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32, (6), pp 1097-1111

[32] Pudil, P., Novovicova, J., Kittler, J.: 'Floating Search Methods in Feature Selection', Pattern Recognition Letters, 1994, 15, (10), pp 1119-1125

[33] Kittler, J., Hatef, M., Duin, R., Matas, J.: 'On Combining Classifiers', IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20, (3), pp 226-239

[34] Fierrez, J., Morales, A., Vera-Rodriguez, R., Camacho, D.: 'Multiple Classifiers in Biometrics. Part 1: Fundamentals and Review', Information Fusion, 2018, 44, pp 57-64

[35] Jain, A., Nandakumar, K., Ross, A.: 'Score Normalization in Multimodal Biometric Systems', Pattern Recognition, 2005, 38, (12), pp 2270-2285

[36] Fierrez-Aguilar, J., Krawczyk, S., Ortega-Garcia, J., Jain, A.K.: 'Fusion of Local and Regional Approaches for On-Line Signature Verification', Proc. Int. Workshop on Biometric Recognition Systems, Beijing, China, 2005, pp 188-196

[37] Alonso-Fernandez, F., Fierrez, J., Ramos, D., Gonzalez-Rodriguez, J.: 'Quality-Based Conditional Processing in Multi-Biometrics: Application to Sensor Interoperability', IEEE Transactions on System, Man, Cybernetics, 2010, Part A, 40, (6), pp 1168-1179

[38] Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Ortega-Garcia, J.: 'Exploring Recurrent Neural Networks for On-Line Handwritten Signature Biometrics', IEEE Access, 2018, 6, pp 5128-5138

[39] Zhang, X., Xie, G., Liu, C., Bengio, Y.: 'End-to-End Online Writer Identification With Recurrent Neural Network': IEEE Transactions on Human-Machine Systems, 2017, 47, pp 103-112

[40] Ha, D., Eck, D.: 'A Neural Representation of Sketch Drawings', arXiv preprint arXiv:1704.03477, 2017