Offline continuous adaptation of templates for signature identification

Monica Carfagni, Lapo Governi, Matteo Nunziati

***Abstract*—** Automatic on-line signature identification is a procedure which allows a machine to identify a subject among a cohort of individuals by using only the subject's signature. The following paper deals with features and models required in order to allow a machine to learn and discriminate people on the basis of such a biometric trait. The proposed solution presents a neural network based framework for template adaptation which has demonstrated to improve the resilience of a system, when it has to face with forgeries, that is, fake signatures which are used in order to attack the system and grant unauthorized access to services. The proposed framework provides an improved security level of 35% with respect to non adapted systems.

***Keywords*—** adaptation, artificial neural networks, identification, signature.

I. Introduction

AUTOMATIC signature recognition has been investigated by several authors (as instance [1], [2]) in order to allow machines to verify an user from its own biometric traits. The usefulness of this approach can be envisaged, as instance, in e-commerce and remote transaction authorization [3]. Among the different biometric traits, signature is defined as a behavioral one, that is, a subject's specific trait acquired during life rather than intrinsic to the human biology itself. Nonetheless, it is really common to apply a signature, especially in those activities related to commercial and financial transactions, for that reason signature is widely accepted as a biometric recognition tool [4], despite its non biological nature. Additionally, acquiring technologies are non invasive and human beings can interface with them with ease. Moving from these considerations, it can be understood why the international community has spent time in order to achieve reliable results with signature based recognition systems.

Based on the authors' work published in [5], this paper is aimed to describe an adaptive system able to identify a subject by means of subject's signature, that is, identify a subject without any other information rather than the signature itself (e.g. no ID has to be provided to the machine). Signature recognition and identification methods can be split in two main fields: on-line [6] and off-line [7]. The former involves the usage of an acquiring device able to track the pen movement during the signature (e.g. a digitizing tablet), the latter investigates ways for signature recognition which are based on static signature images (e.g. forensic signature recognition is based on the analysis of signatures made on paper and acquired by scanners). In other terms the on-line family of methods manage a signature as the trajectory of an object which changes its position in time: this paper focuses on such kind of methods.

The reminder of this paper is organized as follow. Section II formulates the identification problem, distinguishing it from the recognition problem; in this section previous improvements to signature recognition by the authors are recalled, which the following paper is based on. Section III presents the adaptation problem and the proposed ANN based solution. Section IV deals with the experimental setup and the results obtained by using the proposed adaptation framework. Eventually, section V presents the conclusions derived from this paper.

II. Identification Problem Formulation

A. Recognition/verification vs identification problem

Biometric techniques involve a two stage procedure for a system to be able to associate an ID to a subject. As first, specific parameters, the features, are extracted from a signature, later, a statistical model is enrolled against such features and stored in a specific facility (a central server rather that an id or credit card) along with an used ID. The model itself, also referred as a template, is used in order to provide a representation of signature statistical properties, which are expected to contain all the relevant information required to detect a subject among a cohort of people. A second step is used for the *recognition* itself. When the subject performs a transaction an ID and a new signature are requested, features are extracted from this new data stream and compared with the model associated to the given ID.

Identification, is a different kind of problem. Given a set of models with an associated ID, only a signature is requested during any new transaction. The features extracted from this signature are then compared to all available templates. The template with the highest score (if this score is major than a fixed threshold θ) is considered the winner template and the associated ID as the one belonging to the subject which requires the transaction. In other terms the machine is able to automatically identify the user among a cohort of individuals, without the usage of any claimed identity (an ID).

It is common to employ log-score ratios in order to link a data stream to an identity. Let $F=\{f_i with \ \ i=1,\dots,n\}$ be the feature set, with i indicating the i-th sample acquired by a

Manuscript received June 10, 2011: Revised version received March 4, 2008. This work was supported by the Tuscany Region POR FSE 2007-2013 Objective 2 Research Plan.

M. Carfagni is with the Dipeartimento di Meccanica e Tecnologia Industriale of University of Florence (monica.carfagni@unifi.it).

L. Governi is with the Dipeartimento di Meccanica e Tecnologia Industriale of University of Florence (lapo.governi@unifi.it).

M. Nunziati is with the Dipeartimento di Meccanica e Tecnologia Industriale of University of Florence (matteo.nunziati@unifi.it).

device at a fixed sampling frequency (usually less than 100 Hz), and Θ_0 be a template, a similarity score S_0 is defined as:

$$S_0 = \frac{1}{n}\sum_{i=1}^{n} P(f_i|\Theta_0) \quad (1)$$

where P is the probability operator. If an alternative template Θ_1 exists, it is possible to estimate a second score S_1 and retrieve the normalized log-score (NLS) as:

$$NLS = \log\left(\frac{S_0}{S_1}\right) \quad (2)$$

Fixed an acceptance threshold θ, if NLS≥θ a subject is considered as the target of Θ_0, that is he/she is the subject the template has been derived from. Otherwise, the subject is considered as the target of Θ_1.

It is worth the trouble to provide a deeper explanation for both Θ_1 and θ. Usually Θ_1 is named alternative model or Universal Background Model (UBM) and it is generated by pooling together feature sets obtained from a reference database R. This model is expected to provide a good estimation of the probability that certain features can occur among different subjects. In other terms it accounts for the typicality of a certain feature. By comparing the score obtained against Θ_0 with the one obtained against an UBM, it is possible to compute the ratio between the probability that a certain feature set is specific of a given target or, rather, it is common among a certain population and, thus, do not provide any real hint about the unknown identity.

The acceptance threshold θ is usually fixed empirically during a test session. In facts, generally, another set of signatures is employed for the tuning of a biometric system: a fictitious set T, which contains at least two sample signatures per subject. One of the signatures (at least) is employed to build up a template associated with targets. Remaining signatures from T are compared with each template in order infer system reliability. Indeed, statistical systems intrinsically induce false acceptance (FA) or false rejection (FR) errors, that is, some targets are wrongly rejected while some impostors are considered as targets. The acceptance threshold is fixed on an application basis, common values being $\theta_{FA=0.01}$, $\theta_{FA=0.1}$ or θ_{EER}. We define here $\theta_{FA=0.01}$ and $\theta_{FA=0.1}$ as the thresholds which allow the system to produce respectively at most an FA of 0.01% and 0.1%, while θ_{EER} is the value for which an Equal Error Rate (EER) is attained, that is FA equates FR.

B. Feature sets

On-line signature recognition/identification requires the employment of digitizing tablets. Such tools allow to record several temporal patters, such as: the pen position on the tablet (x,y), its pressure (p) and the azimuth and altitude angles of the pen with respect to the tablet (γ,φ). Dealing with signatures implies three degrees of freedom in the acquired data: the same signature, reproduced in different sessions, can laid at any place on the tablet and can be produced with different orientations. This implies that the (x,y) pair for each sampled point can change session by session.

In order to remove such kind of session dependent variability, a standard rototranslation is performed before any feature is computed from raw data. As first, the center of mass (x_0,y_0) of the signature is computed by:

$$x_0 = \frac{1}{n}\sum_{i=1}^{n} x_i \quad y_0 = \frac{1}{n}\sum_{i=1}^{n} y_i \quad (3)$$

than the average angle with respect to the tablet coordinate system is estimated with:

$$\beta = \frac{1}{n}\sum_{i=1}^{n} \arctan \dot{y}_i / \dot{x}_i \quad (4)$$

where the $\dot{(\cdot)}$ operator identifies the first order derivative over time. A numerically robust computation is [1]:

$$\dot{q}_i = \frac{1}{2}\cdot\sum_{\tau=1}^{2} \tau\cdot(q_{i+\tau} - q_{i-\tau}) / \sum_{\tau=1}^{2} \tau^2 \quad (5)$$

where q is a generic variable. Eventually, the whole data set is rotated and translated so that the new β will be null and the center of mass will fit the origin of the reference system.

It is quite common to not rely on raw data for biometric recognition, rather features are extracted by reworking raw input in a proper manner: as instance both in speaker and face recognition spectral features are employed. State-of-the-art signature features involve the following derived measures: the trajectory tangent angles δ, the instantaneous velocities v. Previous work on this topic [1] has demonstrated the scarce usefulness of the pair (γ,φ), pointing out how the remaining variables can better encoding subject specific traits. The resultant feature vector is thus w = [x, y, p, δ, v] , where δ and v are computed as:

$$\delta = \arctan \dot{y}_i / \dot{x}_i \quad v = \sqrt{\dot{x}^2 + \dot{y}^2} \quad (6)$$

In addition to the base feature vector w, delta features between adjacent frames are computed as $\Delta w = \dot{w}$, and delta delta as $\Delta\Delta w = \dot{\Delta} w$.

This approach has empirically demonstrated an increased discrimination capability in any field of biometry and has been recently motivated at theoretical level too [8]. Therefore, the final 1x15 feature vector is f = [w, Δw, ΔΔw] = [x, y, p, δ, v, $\dot{x}$, $\dot{y}$, $\dot{p}$, $\dot{\delta}$, $\dot{v}$, $\ddot{x}$, $\ddot{y}$, $\ddot{p}$, $\ddot{\delta}$, $\ddot{v}$]. Eventually, features are normalized so that each component of f is mapped to a canonical normal distribution with zero mean and unitary variance.

In [5], the authors have reviewed the signature process from a physical perspective, which is compatible with the general theoretical framework worked out in [8]. Briefly, the whole act of signature making can be reduced to the motion of a point in space (the pen tip); therefore, the signature can be described by the classical problem of a material point moving in a bi-dimensional space (in this work we have leaved the pressure out of this model). According to classical equations of mechanics, a material point moving on a straight path can be

represented by a dynamic system, where the state is defined by the vector (x, y, $\dot{x}$, $\dot{y}$), that is, point's position and instantaneous velocity, while the input is defined by the acceleration provided to it by external forces: ($\ddot{x}$, $\ddot{y}$) - ($\ddot{}$) being the second order derivative over time. Generalizing this model to a point moving on a generic path, centripetal acceleration $\ddot{\delta}$ comes as additional input and the point's state can be expressed by a generalized vector such as: (x, y, δ, $\dot{x}$, $\dot{y}$, $\dot{\delta}$), where the added parameters account for the instantaneous tangent angle and angular velocity. Moving from this model and by adding the pressure information, authors proposed the following reduced 1x10 feature vector:

$$f' = [x, y, \delta, p, v, \dot{\delta}, \dot{p}, \dot{v}, \ddot{\delta}, \ddot{p}] \quad (7)$$

where $\dot{v} = \sqrt{\ddot{x}^2 + \ddot{y}^2}$, and other derivatives are computed according to eq. (7).

This approach has demonstrated to improve both the discriminatory capability of a biometric system as well as its resilience to fake signatures (forgeries).

C. State-of-the-art models

In order to compute a proper template, stochastic or statistic models are employed. The best performing stochastic model has been shown to be the Hidden Markov Model (HMM) [1]. Such a model being extremely complex and slow to be computed, studies have been carried out in [2] demonstrating that the same performance can be attained by avoiding any temporal information about the signature patter, that is, by building a statistic template. A commonly employed statistic model is the Gaussian Mixture Model (GMM) [2]. Given a random variable x and number k of multivariate normal distributions $N(x,\mu_j,\Sigma_j)$, a GMM based template is defined as:

$$\Theta = \sum_{j=1}^{k} \alpha_j N(x, \mu_j, \Sigma_j) \quad \text{with } \alpha_j \in \mathbb{R}, \forall j \text{ constrained to } \sum_{j=1}^{k} \alpha_j = 1 \quad (8)$$

with such a formulation, each term of the sum on the right side of eq. (1) is computed as:

$$P(f_i|\Theta) = \sum_{j=1}^{k} \alpha_j N(f_i, \mu_j, \Sigma_j) \quad (9)$$

The generation of a GMM template involves the computation of the unbiased estimators for each mean μ_j and covariance matrix Σ_j (which is usually constrained to be diagonal) as well as the weights α_j. This estimation has not closed form solution, therefore the well known iterative Expectation-Maximization (EM) algorithm [9] is employed for the task.

Concerning the model employed in this paper, other fields of biometry make wide use of the so named UBM-GMM model. This model has been introduced in speaker recognition in [10] and represents a special case of the Maximum A Posteriori (MAP) estimator for HMM parameters, described in [11] and extended in [12]. In brief, the classical EM algorithm needs a relevant number of data for its estimates to be accurate enough for a recognition. As a matter of fact, common biometric traits do not provide such an amount of data and the overall system accuracy is degraded by this lack. By applying MAP estimation to biometric data, authors of [10] have made their system less sensitive to this issue. The procedure, detailed in [10] and [11], can be synthesized as follow: EM is applied to compute an UBM model - which does not suffer of data lack, being generated by pooled data -, then the MAP algorithm is applied in order to derive templates from subject's features.

The MAP algorithm interpolates between the UBM parameters and the template parameters as computed by directly applying EM to the subject's features. Specifically, the MAP procedure interpolates at each iteration of the EM algorithm. According to terms defined in eq. (3), template parameters are estimated iteratively as:

$$\begin{cases} \mu_j^+ = \mu_{UBM} + D_\mu \mu_j^- \\ \Sigma_j^+ = \Sigma_{UBM} + D_\Sigma \Sigma_j^- \\ \alpha_j^+ = \alpha_{UBM} + D_\alpha \alpha_j^- \end{cases} \quad (10)$$

where j accounts for the iterations of the EM algorithm and $D_{(.)}$ are diagonal relevance matrices. Each entry of $D_{(.)}$ defines a weight to be applied in the sum. Possible values for $D_{(.)}$ are proposed in [10] and [12]; namely in [10] an a priori set of weights is employed, while in [12] a more advanced adaptive method is presented.

The authors in [5] have proposed the a priori version of MAP for signature recognition tasks, demonstrating that this approach increases, again, the resilience of a system to forgery-based attacks. According to previous results in recognition, the templates used in this paper are based on the same computational model.

III. ADAPTATION PROBLEM FORMULATION

A. Continuous vs discontinuous adaptation strategies

Among the several frameworks for verification and identification, two families of procedures can be defined: static templates and adapted templates. Static templates are those which are created during the enrolling phase, as described in section II.B, and remain unchanged during the whole life-cycle of a registered ID; adapted templates are those which are created during the enrollment phase and then updated during their usage.

The idea behind adaptation is to overcame the lack of data which afflicts biometric applications [13]. Similarly to the MAP approach, adaptation is employed in order to reuse the data stream coming from each transaction. As a matter of fact, the biggest problem to face with in adaptation is the real ID of a data stream: if the enrolling phase can assure the ID of a data stream, any data stream coming from a transaction is identified by means of the biometric engine. As previously stated, each biometric engine works in the statistic filed, that is, the outcome of a biometric system is not 100% reliable due to the possibility of FA and FR occurrences. For that reason, a solution has to be found in order to define the degree of

confidence about a certain ID, as such an ID defines how and when a data stream can be used to update a given template. Classical, or *discontinuous*, adaptation relies on the fact that the winner ID is considered as the source of a data stream, and it is adapted according to the incoming data stream. As described in [13], adaptation can be performed in several ways:

1. the UBM is retrained with MAP by using the new data stream, old template is discarded and substituted by the new one;
2. the template is used as base model and MAP is applied to it in order to adjust it according to the new data stream;
3. the original data stream is retained and the new incoming streams are added to it, the new augmented data set is used to adapt the UBM via MAP, discarding the old template.

The biggest problem in such a framework is related to the acceptance threshold θ, which needs a fine tuning and tends to diverge from optimality after a number of adaptations [13]. Reviewing this approach from a probabilistic point of view, the method can be considered as follow:

algorithm A:

a) each template of a set Ξ of t models is labeled with a confidence level of 0 or 1;
b) being t^* the winner template, a confidence level of 1 is assigned to t^*;
c) remaining templates are associated to a null confidence level;
d) only the template labeled with 1 is adapted.

In other terms, adaptation is discontinuous for each template, as they are updated if and only if they receive a label of 1. In order to overcome this limitation, the concept of *continuous* adaptation is described in [13]. A so named WMAP approach is described for the assessment of a data stream ID. The WMAP approach consists in computing a continuous confidence level which belongs to the range [0,1]; similarly to the previous algorithm, the general adaptation procedure is described as:

algorithm B:

a) each template of a set Ξ of t models is labeled with a confidence level ranging from 0 to 1;
b) all templates are adapted by applying a relevance matrix $D_{(.),t}$ proportional to the received label.

According to the modified method, all templates are adapted at each iteration, producing a continuous update, whose rate is defined by the confidence level, which associates each data stream to each registered template. Among the possible ways to compute such a confidence label and the related weight, in this paper (section II.F) authors present a novel method based on ANN, which reflects the general schema of WMAP, but differs from it in the implementation details.

B. On-line vs off-line adaptation

Template adaptation can be performed either on-line or offline as described in [14]. On-line adaptation can be used if several samples are available during the identification task. As instance face identification can be performed by using a video stream, faces can be extracted by several frames of the video and the incoming data stream can be used to accumulate information as per 3. . In such a way the identification task provides a time varying solution which is updated in a quasi-real-time manner. The final decision is kept by the biometric engine when such solution (the expected ID) becomes stable with respect to time.

Off-line adaptation is employed when only one feature set per transaction is available. It deals with the retraining of a certain model after a given transaction is terminated. In other terms, a set of templates is used to identify a subject and, later, the templates (or just the most fitting one) are updated on the basis of the income data stream and its associated ID. The updates models will be employed in the next incoming transaction. In the latter case, adaptation is used to improve the accuracy session after session, therefore the benefits of this approach are related to the *inter*-session accuracy of the engine and not to the *intra*-session reliability.

This paper deals with off-line adaptation. Moreover, as raw data are a risk with respect to privacy issues (as they contain sensible data of an user), we present here an approach based on template adaptation as per 2. .

C. Proposed framework for confidence estimation

Regardless of the off-line/on-line method, the biggest issue in template adaptation is related to the definition of a confidence level about a certain ID. In other terms, the system has to understand which is the probability that a certain data stream belongs to a given template.

In order to solve this problem, Artificial Neural Networks (ANN) are employed in this paper. ANN are trained according to the procedure described in [15]. The input of the network is the NLS obtained by testing the data stream against each registered template; the output is expected to be a well-calibrated loglikelihood ratio (LLR) as in [15]. According to the Bayesian theorem, LLRs are related to the probability of template matching by:

$$\log\left(\frac{P_{match}}{1-P_{match}}\right)=LLR+\log\left(\frac{P_{prior}}{1-P_{prior}}\right) \quad (11)$$

where P_{prior} is a prior probability, here fixed to 0.5. Moving from equation (9), it is possible to compute P_{match} ad assign a confidence label to each comparison between a data stream and a template. This probability is used to decrease the a priori weight of the data stream, which is employed during the MAP algorithm.

Namely, being t the number of registered templates in a set Ξ and being $D_{(.)}$ anyone of the relevance matrices, as per eq. (10), the following steps are applied:

algorithm C:

a) each NLS is processed by the ANN obtaining a confidence label ranging from 0 to 1;

b) each template of a set Ξ of t models is labeled according to the related confidence level;
c) a new relevance matrix set $\tilde{D}_{(.),t} = P_{match} D_{(.),t}$ is computed at the end of each session, for each template in Ξ;
d) a MAP adaptation is performed against the templates by using matrices $\tilde{D}_{(.),t}$.

In other terms, the whole set Ξ is updated by using a specific confidence value for each template, in a manner similar to [13]. The direct consequence of this approach is that, if a template attains a confidence of 1 (that is, it is 100% sure that a certain ID is the source of a data stream), it is updated by using the same weighting matrix employed during the enrollment; otherwise the weight of the incoming data stream decreases down to 0 - the template is not updated at all - if the confidence label decreases down to 0 (that is, it is 100% sure that the template is not the source of a data stream).
Moving from the last sentence, it is understandable how to attain an unbiased estimation of LLR is of primary importance in the proposed framework. According to [15], this can be attained if the feed-forward neural network is trained by minimizing the following cost function:

$$Cllr = -\frac{1}{N_T}\sum_{i=1}^{N_T}\log_2\left(1+\frac{1}{\exp(LLR_i)}\right) - \frac{1}{N_I}\sum_{j=1}^{N_I}\log_2\left(1+\exp(LLR_j)\right) \quad (12)$$

where C_{llr} is the name of the cost function as proposed in [15], N_T is the number of target comparisons, that is the comparisons made between a data stream and a template belonging to the same subject, while N_I is the number of impostor comparisons made by testing signatures against the registered templates of other users in Ξ.
The original ANN proposed by [15] was composed by a 2-layer network with a single exponential neuron in the hidden layer and a linear neuron as output. In this work we propose a modified topology, which employes a 2-layer ANN with hyperbolic tangent neurons in the hidden layer and one linear neuron as output. The number of hidden units is defined automatically according to [16]. Briefly the optimal number of hidden neurons h_{opt} is computed as:

$$h_{opt} = \begin{cases} S/d & if\ S/d \le 30 \\ \sqrt{S/(d\cdot logN)} & ,otherwise \end{cases} \quad (13)$$

where S is the total number of samples (N_I+N_T), and d is the dimension of the input space. The employed learning algorithm is the well known stochastic gradient ascend method (SGA).
In order to train the ANN a preliminary recognition session is performed on an auxiliary data set. In this set each subject is associated to a template in Ξ, and cross comparisons between different subjects are performed as in [5]. The obtained scores are associated to a reference label L_R; L_R can assume value {I, T} depending on the kind of comparison: if the data stream belongs to the subject from which the template has been derived, the T label is assigned, otherwise I is used. After the comparison, NLS are processed by the ANN and the output is used in order to compute eq.(21), according to the assigned labels. The ANN coefficients are iteratively updated by the SGA algorithm in order to minimize the cost of the obtained outcomes. Once the ANN has been trained, it is employed for an identification session as per algorithm C.

IV. Experimental setup and results

A. Data-sets and models

In order to test our hypotheses, the myIDea database [17] has been employed. The data set is composed by 3537 signatures collected from 73 different subjects. Each subject has been acquired in different sessions, collecting up to five genuine signatures per session and up to five forgeries and skilled forgeries (the latter made after a period of training, in order to allow the subject to produce a more accurate fake signature). 1173 of these signatures have been employed to train the UBM model. Remaining signatures have been collected in two separate pairs of sets (T_c,I_c) and (T_v,I_v), employed to simulate respectively different claimed identities ($T_{(.)}$) and a set of impostors ($I_{(.)}$) in both calibration and validation. Additionally, for each subject in $T_{(.)}$, a separate database F of related forgeries has been derived, picking fake signatures generated by the other subjects.

During the test, both system accuracy in signature discrimination (set $T_{(.)}$ vs set $I_{(.)}$) and system sensitivity to forgeries (T_v vs. F) have been evaluated. Result are reported in terms of DET plots [18] for the T_c vs. I_c test, FR and FA have been employed to evaluate the validation test, while FA is used for the evaluation of the T_v vs. F test.

Being literature results based on different data sets, a baseline model has been built. The baseline templates are obtained by MAP training GMM against the improved 1x10 feature vectors described in section II.B. A second model SYS1 is composed by the same type of templates but off-line adaptation has been employed to retrain the models during simulated transactions. It is common to reduce the MAP procedure so that only means are adapted, that is the UBM and the templates share the same Σ and α. This procedure has shown to provide very good results (compare [10] as instance) and has been applied also in this paper.

The rationale of the presented test is the following: the baseline system will provide a reference performance; by comparing SYS1 with the baseline it will be possible to assess the performance effects induced by the off-line adaptation.

B. System training and calibration

In [2] has been shown how model size predictors such as the Minimum Description Length do not apply to diagonally constrained GMM. As a matter of fact the correct number of components k has to be defined iteratively on an a posteriori analysis of EER [2].

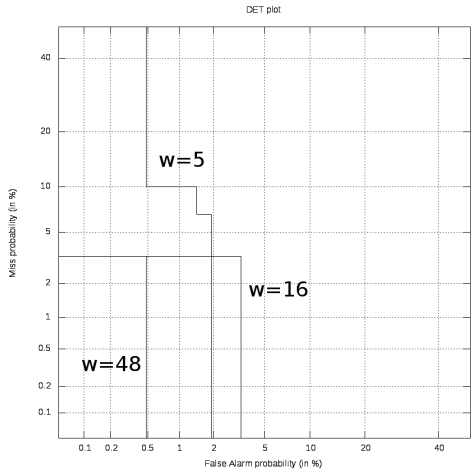

Fig. 1- DET plot for the baseline system – 512 components.

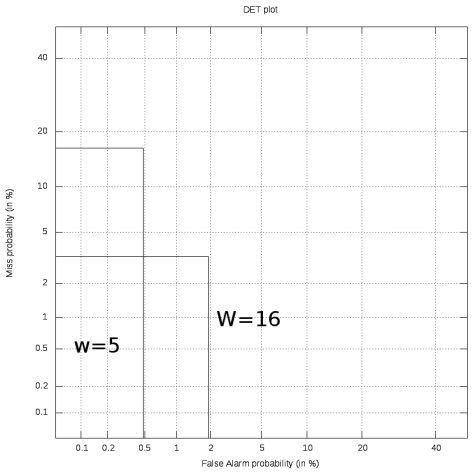

Fig. 4 - DET plot for SYS1 – 512 components.

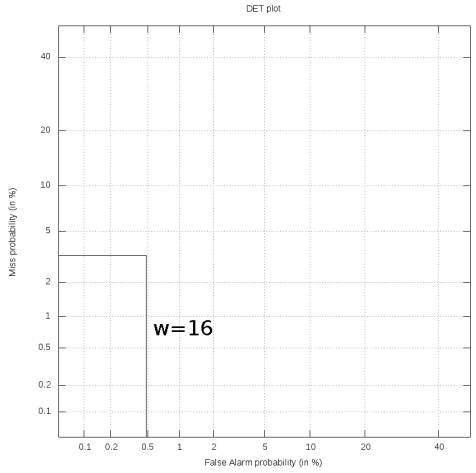

Fig. 2- DET plot for the baseline system – 1024 components.

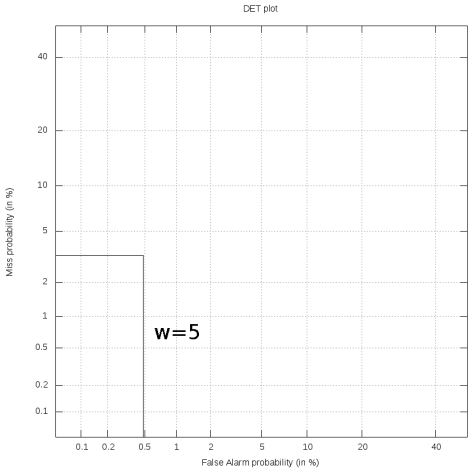

Fig. 5 - DET plot for SYS1 – 1024 components.

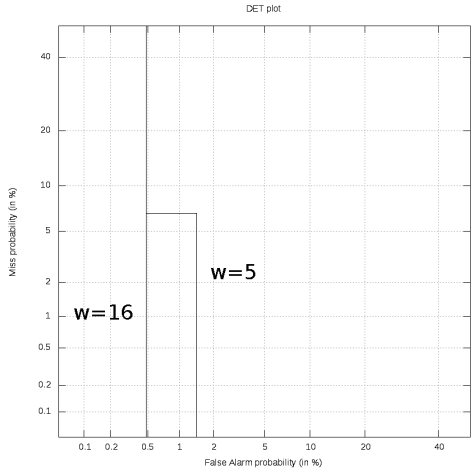

Fig. 3- DET plot for the baseline system – 2048 components.

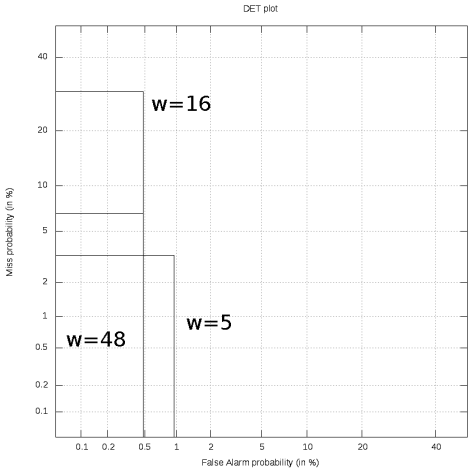

Fig. 6 - DET plot for SYS1 – 1024 components.

In this paragraph we report the different levels of equal error rates for the standard and adapted models, providing the evidence in support of the topologies used in the following paragraphs. In the figures 1 to 3, different DET curves are reported for the baseline system with size 512, 1024 and 2048 and relevance matrix with a priori weight w of 48, 16 and 5. Similarly, DET plots are presented for the off-line adaptation employed in SYS1 (Fig. 4 to 6).

The most relevant aspect of the presented plots is their limited fragmentation. Even if 2390 comparisons have been accomplished during tests, the curves remain monolithic due to the low level of FA attained by the systems. In some plots only one curve is show: this is due to the absence of FA and FR during calibration.

According to the procedure described in section III.C, an auxiliary recognition session has been carried out, which has lead to an ANN with 17 hidden units. Details about the recognition session can be derived from [5]. Eventually, the plots have been employed in order to derive the optimal $\theta_{EER,sys}$ for each system.

C. System validation

Table 1 reports the results attained during the validation phase and the resilience test. Accuracy is computed according to FA and FR and $HTER=(FA+FR)/2$ is also reported as synthetic indicator (HTER stays for Half Total Error Rate).

Tab. 1- validation results

System	size	Weight	HTER [%]	FR [%]	FA [%]
baseline	512	5	7.84	15.69	0
baseline	512	16	**1.19**	**1.96**	**0.41**
baseline	512	48	2.58	3.92	1.24
baseline	1024	5	2.94	5.88	0
baseline	1024	16	1.96	3.92	0
baseline	1024	48	3.92	7.84	0
baseline	2048	5	3.92	7.84	0
baseline	2048	16	3.92	7.84	0
baseline	2048	48	2.94	5.88	0
SYS1	512	5	*1.24*	*0*	*2.48*
SYS1	512	16	4.90	9.80	0
SYS1	512	48	9.80	19.60	0
SYS1	1024	5	**1.29**	**1.96**	**0.62**
SYS1	***1024***	***16***	***3.92***	***7.84***	***0***
SYS1	1024	48	8.82	17.65	0
SYS1	2048	5	1.96	3.92	0
SYS1	2048	16	3.92	7.84	0
SYS1	2048	48	6.86	13.72	0

During the validation phase, a system is tested as in a production environment, therefore, it is expected that the outcome of a biometric comparison will be a binary value (0,1). Such a value should be 1 for the expected ID and 0 for all the other registered IDs. In this test, if an impostor tries to access the system, he is expected to force the system with his own signature, that is, no forgeries are proposed here.

The decision about the ID (register or not) is kept according to the method described in section II.A. If the winner template t^* is not the real ID related to the signature an FA event is raised; on the opposite, if an ID is not recognized as the author of the signature a FR event is raised. The winner template has to obtain a score higher that the selected $\theta_{EER,sys}$.

Table 2 reports the results obtained during the resilience test, that is, the test performed injecting forgeries in the system. This kind of tests is aimed to define the resistance of the biometric system, when attacked with fake signatures. In this case only impostor IDs are presented to the system, therefore, an optimal engine should reject any tried access. Due to this configuration only FA errors can arise.

Tab.2- resilience results

System	size	Weight	FA [%]
baseline	512	5	4.09
baseline	512	16	7.73
baseline	512	48	7.73
baseline	1024	5	3.64
baseline	1024	16	5.91
baseline	1024	48	3.64
baseline	2048	5	**0.46**
baseline	2048	16	1.82
baseline	2048	48	5.00
SYS1	512	5	12.27
SYS1	512	16	**0**
SYS1	512	48	**0**
SYS1	1024	5	9.10
SYS1	***1024***	***16***	***0***
SYS1	1024	48	**0**
SYS1	2048	5	4.55
SYS1	2048	16	**0**
SYS1	2048	48	**0**

D. Results discussion

According to results presented in Tab.1, both the baseline system and the SYS1 configurations can attain the same reliability level. Considering the synthetic indicator HTER, the most reliable systems appear to be the baseline with 512 components and an a priori weight of 16, the SYS1 with both 512 and 1024 components and a weight of 5. Looking at the specific rates of FA and FR it emerges that SYS1 with 512 components induces relevant FA errors: even if the HTER reaches really good levels, such an unbalanced behavior can induce a number of unauthorized accesses to biometrically enabled facilities. For that reason such a system has to be rejected as too permissive.

Analyzing Tab.2, the adaptation strategy appears to boost the resilience of the system: the baseline attains an average FA of 4.46%, while the average FA for SYS1 is 2.88% (-35%). One relevant item is the instability of the procedure, which, if

adequately initialized, can provide a virtually perfect rejection of all forgeries. On the opposite, if a bad initialization is accomplished (due to badly identified IDs), FA can increase up to 60% more than the baseline (12.27% vs. 7.73%). this effect is remarkably evident in SYS1 with 512 components and weight 5, where a relevant amount of FA is produced.

Considering the merged results and the need of a secure system, the most balanced and robust solution is still accomplished via adaptation: the best overall result is attained with SYS1 by using 1024 components and a weight of 16. In this case FA is always 0% even if under attack, while the FR remains under 8%. A similar result can be attained without adaptation, nonetheless an FA of at least 0.46% is present under attack and the template size must be doubled (from 1024 to 2048).

V. Conclusion

The following paper deals with features and models for on-line signature identification. The proposed solution approaches the signature making process as the motion of a point in a bi-dimensional space and models its statistic properties via the well known MAP training of GMM templates. In order to harden the system with respect to possible attacks made by means of forgeries, an ANN based adaptation procedure has been presented.

Comparing our approach to non adapted systems an improved average robustness to attacks of 35% has been attained. Namely, during experiments has been possible to halve the size of users templates, while attaining a 0% FA system. One relevant aspect which deserves attention and further investigation is the instability of the adaptation process: if a system is not correctly trained, adaptation can lead the engine to the drift, causing major issues with false acceptances.

References

[1] J. Fierrez, J. Ortega-Garcia, D. Ramos and J. Gonzalez-Rodriguez, "HMM-based on-line signature verification: feature extraction and signature modeling", *Pattern Recognition Letters*, Vol. 28, n. 16, pp. 2325-2334, 2007.

[2] J. Richiardi, A. Drygajlo, "Gaussian Mixture Models for On-line Signature Verification", *Proceedings of the 2003 ACM SIGMM workshop on Biometrics*, 2003.

[3] O. Horak, "Web-application User Identification", *Proceedings of the International Conference on Applied Computer Science*, pp. 414-418, 2010

[4] F.-M.E. Uzoka, T. Ndzinge, "Empirical analysis of biometric technology adoption and acceptance in Botswana", *The Journal of Systems and Software,* No.82, pp. 1550–1564, 2009.

[5] M. Carfagni, M. Nunziati , "An Improved Model and Feature Set for Signature Recognition ", *Proceedings of the International Conference on COMPUTERS and COMPUTING*, vol .1, pp.75-80, May 2011.

[6] S. A. Rahman, R. Yusof, S. Mutalib, M. Yusoff, A. Mohamed , "Slant Algorithm for Online Signature Recognition", *Proceedings of the 10th WSEAS International Conference on EVOLUTIONARY COMPUTING*, pp.152-157, 2009.

[7] B. Kovari, H. Charaf, "Statistical Analysis of Signature Features with Respect to Applicability in Off-line Signature Verification ", 14th *WSEAS International Conference on COMPUTERS,* pp.473-478, 2010

[8] J. Richiardi , K. Kryszczuk , and A. Drygajlo, "Static models of derivative-coordinates phase spaces for multivariate time series classification: an application to signature verification", *ICB '09 Proceedings of the Third International Conference on Advances in Biometrics*, pp. 1200–1208, 2009.

[9] A.P. Dempster, N.M. Laird, D.B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm", *Journal of the Royal Statistical Society*, Series B (Methodological) Vol.39, No.1, pp. 1–38, 1977.

[10] D.A. Reynolds, T.F. Quatieri, R.B. Dunn, "Speaker Verification Using Adapted Gaussian Mixture Models", *Digital Signal Processing*, Vol.10, pp. 19–41, 2000.

[11] J. L. Gauvain, C.-H. Lee, "Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains", *IEEE Trans. Speech Audio Process.*, Vol.2, pp. 291–298, 1994.

[12] P. Kenny, G. Boulianne, P. Dumouchel, "Eigenvoice modeling with sparse training data", *IEEE Trans. Speech Audio Processing*, Vol. 13, No. 3, 2005.

[13] A. Preti, J.F. Bonastre, F. Capman, "A continuous unsupervised adaptation method for speaker verification ",*International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering*, 2006

[14] N. Poh, N. Johansson, C. McCool, "D4.8: Advanced Model Adaptation System", *MOBIO Mobile Biometry European Project Public Report*, 2010.

[15] Brummer N., du Preez J., "Application-independent evaluation of speaker detection", C*omputer speech and language*, vol. 20, pp.230-275, 2006.

[16] S. Xu and L. Chen, " A Novel Approach for Determining the Optimal Number of Hidden Layer Neurons for FNN's and Its Application in Data Mining ", *5th International Conference on Information Technology and Applications*, pp.683-686, 2008.

[17] B. Dumas, C. Pugin, J. Hennebert, D. Petrovska-Delacrétaz, A. Humm, F. Evéquoz, R. Ingold, D. Von Rotz, "MyIdea - Multimodal Biometrics Database, Description of Acquisition Protocols"**,** *proc. of Third COST 275 Workshop*, pp. 59-62, 2005.

[18] Martin A., Doddington G., Kamm T., Ordowski M., Przybocki M., "The DET curve in assessment of detection task performance", *proceedings of the 5th Eurospeech conference*, pp. 1895–1898, 1997.