Zapata-Zapata, Gabriel Jaime; Arias-Londoño, Julián David; Vargas-Bonilla, Jesús
Francisco; Orozco-Arroyave, Juan Rafael
On-line signature verification using Gaussian Mixture Models and small-sample learning
strategies

On-line signature verification using Gaussian Mixture Models and small-sample learning strategies

Verificación de firmas en línea usando modelos de mezcla Gaussianas y estrategias de aprendizaje para conjuntos pequeños de muestras

Gabriel Jaime Zapata-Zapata[1], **Julián David Arias-Londoño**[2*], **Jesús Francisco Vargas-Bonilla**[1], **Juan Rafael Orozco-Arroyave**[1,3]

[1]Departamento de Ingeniería Electrónica y Telecomunicaciones, Universidad de Antioquia. Calle 67 # 53-108. A. A. 1226. Medellín, Colombia.

[2]Departamento de Ingeniería de Sistemas, Universidad de Antioquia. Calle 67 # 53-108. A. A. 1226. Medellín, Colombia.

[3]Pattern Recognition Lab, Friedrich-Alexander University Erlangen-Nürnberg. Martensstraße 3. 91058. Erlangen, Germany.

ABSTRACT: This paper addresses the problem of training on-line signature verification systems when the number of training samples is small, facing the real-world scenario when the number of available signatures per user is limited. The paper evaluates nine different classification strategies based on Gaussian Mixture Models (GMM), and the Universal Background Model (UBM) strategy, which are designed to work under small-sample size conditions. The GMM's learning strategies include the conventional Expectation-Maximisation algorithm and also a Bayesian approach based on variational learning. The signatures are characterised mainly in terms of velocities and accelerations of the users' handwriting patterns. The results show that for a genuine vs. impostor test, the GMM-UBM method is able to keep the accuracy above 93%, even when only 20% of samples are used for training (5 signatures). Moreover, the combination of a full Bayesian UBM and a Support Vector Machine (SVM) (known as GMM-Supervector) is able to achieve 99% of accuracy when the training samples exceed 20. On the other hand, when simulating a real environment where there are not available impostor signatures, once again the combination of a full Bayesian UBM and a SVM, achieve more than 77% of accuracy and a false acceptance rate lower than 3%, using only 20% of the samples for training.

RESUMEN: El artículo aborda el problema de entrenamiento de sistemas de verificación de firmas en línea cuando el número de muestras disponibles para el entrenamiento es bajo, debido a que en la mayoría de situaciones reales el número de firmas disponibles por usuario es muy limitado. El artículo evalúa nueve diferentes estrategias de clasificación basadas en modelos de mezclas de Gaussianas (GMM por sus siglas en inglés) y la estrategia conocida como modelo histórico universal (UBM por sus siglas en inglés), la cual está diseñada con el objetivo de trabajar bajo condiciones de menor número de muestras. Las estrategias de aprendizaje de los GMM incluyen el algoritmo convencional de Esperanza y Maximización, y una aproximación Bayesiana basada en aprendizaje variacional. Las firmas son caracterizadas principalmente en términos de velocidades y aceleraciones de los patrones de escritura a mano de los usuarios. Los resultados muestran que cuando se evalúa el sistema en una configuración genuino vs. impostor, el método GMM-UBM es capaz de mantener una precisión por encima del 93%, incluso en casos en los que únicamente se usa para entrenamiento el 20% de las muestras disponibles (equivalente a 5 firmas), mientras que la combinación de un modelo Bayesiano UBM con una Máquina de Soporte Vectorial (SVM por sus siglas en inglés), modelo conocido como GMM-Supervector, logra un 99% de acierto cuando las muestras de entrenamiento exceden las 20. Por otro lado, cuando se simula un ambiente real en el que no están disponibles muestras impostoras y se usa

* Corresponding author: Julián David Arias Londoño
e-mail: julian.ariasl@udea.edu.co

únicamente el 20% de las muestras para el entrenamiento, una vez más la combinación del modelo UBM Bayesiano y una SVM alcanza más del 77% de acierto, manteniendo una tasa de falsa aceptación inferior al 3%.

1. Introduction

Biometrics measures individuals' unique physical or behavioural characteristics with the aim of recognising or authenticating identity. The most common physical biometrics include fingerprints, hand or palm geometry, retina, iris, or facial characteristics, among others. On the other hand, behavioural characteristics include signature, voice (which also has a physical component), keystroke pattern, and gait, among others. According to [1], signature and voice technologies are one of the most developed. The handwritten signature is recognised as one of the most widely accepted personal attributes for identity verification. The signature is a symbol of consent and authorisation, especially in the credit card and bank-checks environment, and has been an attractive target of fraud for a long time. Currently, there is a growing demand for the processing of individual identification to be faster and more accurate, therefore the design of a robust automatic signature verification system becomes an important challenge.

A comparison of signature verification with other recognition technologies, e. g. fingerprint, face, voice, retina and iris scanning, reveals that signature verification has several advantages as an identity verification mechanism. Firstly, signature analysis can only be applied when the person is/was conscious and willing to write in the usual manner. To give a counter example, a fingerprint may also be used when the person is unconscious, i.e. drugged state. Forging a signature is deemed to be more difficult than forging a fingerprint, given the availability of sophisticated methods [2].

Unfortunately, signature verification is a difficult discrimination problem since a handwritten signature is the result of a complex process depending on the physical and psychological conditions of the signer, as well as the conditions of the signing process [3]. There are two major methods of signature verification. One is an on-line method to measure sequential data, such as handwriting speed and pen pressure, with a special device. The other one is an off-line method that uses an optical scanner to obtain handwriting data written on paper. The dynamic information of the pen-tip (stylus) movement such as pen-tip coordinates, pressure, velocity, acceleration, and pen-up/pen-down, can be captured by a tablet in real time but not by an image scanner [4].

The normal variability of signatures constitutes the greatest obstacle to be met in achieving automatic verification. Signatures vary in their complexity, duration, and vulnerability to forgery [5]. Moreover, signers vary in their coordination and consistency. Problems of signature verification are addressed by taking into account three different types of forgeries [6]: random forgeries, produced without knowing either the name of the signer nor the shape of its signature; simple forgeries, produced knowing the name of the signer but without having an example of his signature; and skilled forgeries, produced by people who attempt to imitate the original signature with prior knowledge of it. Clearly, the problem of signature verification becomes more and more difficult when passing from random to simple and skilled forgeries, the latest being a much more difficult task even more considering that humans use to make errors in several cases. Indeed, exercises in imitating a signature often allow humans to produce forgeries very similar with respect to the originals, making their discrimination practically impossible. In many cases, the distinction is complicated even more by the large variability introduced by some signers when writing their own signatures [4].

Unlike conventional pattern recognition systems, where every object to be recognised/classified is represented as a feature vector, the on-line signature verification (OSV) requires the processing of time dependent signals where every signature produces a set of feature vectors (one per instant of time). Since the observations at the time t are not independent from the previous observations, this problem has been addressed by means of stochastic models able to model that dependence, such as hidden Markov models (HMM) [7]. In most of the cases, the number of available samples for training the system is small, so one of the main challenges of the OSV problem, is to build a system able to achieve high recognition rates with a small number of training samples. This fact implies an important drawback for systems based on HMM because it is well-known that the large number of parameters of a HMM requires a large number of training samples in order to a get a properly fit of the model.

The assumption of independence among the observations is a strategy successfully applied in the speaker verification field. In this case, the verification task can be seen as a multiple-instance learning problem, where, instead of receiving a set of instances which are individually labelled, the system receives a set of labelled bags, where each bag is the set of observations representing a signature. This approach allows the use of models with less computational load and lower requirements on the number of training samples. This work explores the use of different strategies based on a class of generative models called Gaussian Mixture Models (GMM), which have the advantage of being able to process signals with different length, such as the ones coming from on-line digitised signatures, without the need of a previous standardisation of the signal length. This kind of models have been extensively used for speaker recognition [8], and they have also been tested for on-line signature verification [9]. GMMs are conventionally trained using the Expectation-Maximisation (EM) algorithm, which is an implementation of the Maximum Likelihood criterion. The EM algorithm provides a simple and quick way to train a GMM. However, it presents three main drawbacks. First of all, the EM algorithm requires a considerable number

of samples for training the models. Second, it is sensitive to overfitting, i.e., it fits to the training data but lacks the generalisation ability to make accurate predictions for new data, and finally, it does not provide a compact way to estimate the correct number of Gaussian components M, so it must be set using a cross-validation strategy. In order to overcome the first two pointed out limitations, in [8], it was proposed a training strategy called GMM-Universal Background Model (GMM-UBM). It consists in training a "universal" model (a class independent GMM adjusted from all the observations in the universe), and using a Bayesian adaptation procedure to transform the UBM into a class-dependent model. This strategy has proved to achieve better results than the conventional GMM strategy [8, 10]. Following this approach, a system combining the advantage of discriminative and generative models was proposed in [11]. The system used a UBM, and more precisely, the mean vectors of user-dependent adapted GMM, to construct a new feature space where a Support Vector Machine (SVM) was employed to take the final decision. This method is typically called GMM-Supervector (GMM-SVM).

Although the UBM and GMM-SVM approaches have obtained significant improvements in comparison to the standard GMM strategy, in both cases the training of the UBM is still based on the EM algorithm, and its corresponding drawbacks remain. Recently, a full Bayesian learning for GMM has been proposed [12], which transforms the GMM into a hierarchical Bayesian model assigning prior distributions to the GMM parameters. The training algorithm based on this approach is called Variational EM, and in addition to show better performance under small sample size conditions, it provides a semi-automatic way to select the optimum number of Gaussian components, with the consequent reduction of the computational load during the training stage.

This paper addresses the problem of on-line signature verification and evaluates nine different classification strategies based on GMM, including a new variational version of the GMM-SVM model. The main aim of the work is to determine whether the strategies based on UBM, can reduce the requirements on the size of training set, in order to enable verification systems to operate in real situations, i.e. when the number of signatures available per user is quite limited. The paper is organised as follows: section 2 presents the features used to characterise the signatures and the classification methods; section 3 exposes the database and the experimental setup, as well as the results obtained. Finally, section 4 presents some conclusions derived from the results.

2. Methods

2.1. Characterization

The input signals from a digitising tablet include the position in the x and y axes, and the pressure (pr) along the time during the writing of the signature. The position in the x and y axes were used to derive six additional dynamical features as suggested in [7]. The set of features includes:

- $Velocity: v = \sqrt{(\dfrac{dx}{dt})^2 + \dfrac{dy}{dt})^2} = \sqrt{(v_x^2 + v_y^2)}$

  The logarithm of the velocity was also included using
  $$logv = \log(1.0 + v)$$

- $Acceleration: a = \dfrac{dv}{dt}$

- $Tangential\ angle: \phi = \arctan(\dfrac{v_y}{v_x})$

- $Curvature\ radius: c_r = \dfrac{d\phi / dt}{v}$

In a similar way to the velocity, the logPressure defined as $logPr = \log(1.0 + pr)$ was also included. Therefore, for every instant of time t the feature vector $\mathbf{o}_t$ was composed of eleven features as follows (see Eq. (1)):

$$\mathbf{o}_t = [x, y, v_x, v_y, v, logv, a, \phi, c_r, pr, logPr]^T \quad (1)$$

A whole signature is then represented by the complete set of feature vectors observed at different times, which can be expressed as (see Eq. (2)):

$$\mathbf{O} = [\mathbf{o}_1, \mathbf{o}_2, ..., \mathbf{o}_l] \quad (2)$$

Where l is the length of the signature in terms of time instants.

2.2. Modelling and recognition

From a pattern recognition point of view, the classification of the signatures defined as in the Eq. (2), can be understood as a multi-instance learning problem [13]. In this work, this problem is addressed using several strategies based on Gaussian Mixture Models (GMM), with the aim of providing a system able to work under conditions of small number of training samples.

Gaussian Mixture Models - GMM

A GMM is a parametric probability density function represented as a weighted sum of Gaussian component densities. GMMs are commonly used in different tasks as a parametric model of the probability distribution of continuous measurements [8]. Formally a GMM can be expressed as (see Eq. (3)):

$$p(\mathbf{x}\,|\,\Theta) = \sum_{i=1}^{M} w_i \mathcal{N}\left(\mathbf{x}\,|\,\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right) \tag{3}$$

Where $\mathbf{x}$ is a ρ-dimensional continuous-valued data vector (i.e. measurements of features), $w_i, i = 1,...,M$ are the mixtures weights, and $\mathcal{N}\left(\mathbf{x}\,|\,\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right)$ are the Gaussian components. Each component is a ρ-variate multivariate Gaussian function, with mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$. The mixture weights satisfy the constraints $w_i \geq 0$ and $\sum_{i=1}^{M} w_i = 1$. The complete GMM is parametrised by the mean vectors, covariance matrices and mixture weights from all component densities. These parameters are collectively represented by the notation (see Eq. (4)):

$$\Theta = \left\{ w_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i \right\}, \quad i = 1, \ldots, M \tag{4}$$

There are several techniques available for estimating the parameters of a GMM; however, traditionally, the most employed technique is the *maximum likelihood* (ML) estimation [14]. The ML estimation technique finds parameters that maximize the joint likelihood of the training data which are supposed to be independent and identically distributed (iid). Given a set X of N iid observations of ρ features $X = \left\{ \mathbf{x}_1, ..., \mathbf{x}_N \right\}$, the GMM likelihood can be written as (see Eq. (5)):

$$p(X\,|\,\Theta) = \prod_{t=1}^{N} p\left(\mathbf{x}_t\,|\,\Theta\right) \tag{5}$$

Although this expression is a non-linear function of the parameters Θ, the joint likelihood can be maximized with a simple and efficient update procedure called *Expectation-Maximisation* (EM) algorithm [14].

Once the parameters of the GMM were calculated, the detection system is a straight-forward generative classifier. For each class to be recognised (genuine or impostor), the parameters of a different GMM are estimated (Θ_g and Θ_i). Thus, the evaluation is carried out calculating a likelihood ratio, in which for each GMM the a posteriori probability of a particular feature sequence $\mathbf{O} = \{o_1, ..., o_l\}$ (extracted from a particular signature) is estimated. Applying the Bayes' rule and discarding constant prior probabilities, the likelihood ratio in the log domain becomes (see Eq. (6)) [15]:

$$L(\mathbf{O}) = \log p\left(\mathbf{O}\,|\,\Theta_g\right) - \log p\left(\mathbf{O}\,|\,\Theta_i\right) \tag{6}$$

The likelihood ratio is compared with respect to a threshold λ in order to take a decision, accept or reject the signature. In the biometric verification research fields, typically the threshold λ is set to the Equal Error rate threshold [16].

It is worth emphasising that, since the OSV task corresponds to a multi-instance learning, the terms in the log likelihood

ratio must be computed as (see Eq. (7)):

$$\log p(\mathbf{O}\,|\,\Theta) = \frac{1}{l} \sum_{t=1}^{l} \log p\left(\mathbf{o}_t\,|\,\Theta\right) \tag{7}$$

where the $\frac{1}{l}$ scale factor is used to normalise the likelihood with respect to the duration of the signature, avoiding a possible bias due to different pattern lengths of correct and impostor signers. Note that l is the number of feature vector composing a single signature, while N in Eq. (5) is the total number of feature vectors per class (including multiple signatures).

Universal Background Model

The GMM-Universal Background Model (GMM-UBM) is a training strategy where all the available samples are used for training a "universal" model (a conventional GMM), and the class-dependent models are adapted from the UBM. There are several adaptation procedures proposed in the literature, but the most widely used are:

- Maximum a posteriori (MAP): This adaptation maximises the *a posteriori* distribution of the adaptation data $\mathbf{O}$ given the a priori model parameters Θ using the Bayes formula (see Eq. (8)) [17]:

$$\arg\max_{\Theta} p(\mathbf{O}\,|\,\Theta) p(\Theta) \tag{8}$$

Where $p(\mathbf{O}|\Theta)$ is the likelihood function of $\mathbf{O}$ given the model parameters. The adaptation can be performed on all of the parameters of the model, even though in some applications it has been found that the most important parameters to be adapted are the mean vectors [8]. MAP assumes the prior distribution for the mean vectors as Gaussian. The adaptation rules are derived using the EM algorithm, which balances the new estimates on the adaptation data and the prior knowledge. For the mean vectors the adaptation is performed according to (see Eq. (9)):

$$\hat{\boldsymbol{\mu}}_i = \gamma_i E_i\{\mathbf{O}\} + (1 - \gamma_i)\boldsymbol{\mu}_i \tag{9}$$

where $\hat{\boldsymbol{\mu}}_i$ is the adapted mean vector for the component i, $E_i\{\mathbf{O}\}$ is the expected mean feature vector for the adaptation data, and γ_i is the adaptation factor that controls the balance between the new data and prior knowledge. It can be estimated as (see Eq. (10)) [8]:

$$\gamma_i = \frac{n_i}{n_i + \tau} \tag{10}$$

where τ is a fixed relevant factor. n_i is similar to the cumulated responsibility of the component i in the

generation of the new data $\mathbf{O}$, i.e. the E step of the EM algorithm (see Eqs. (11) and (12)). Formally

$$n_i = \sum_{t=1}^{l} p(i \mid \mathbf{o}_t) \qquad (11)$$

where

$$p(i \mid \mathbf{o}_t) = \frac{w_i \mathcal{N}\left(\mathbf{o}_t \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right)}{\sum_{j=1}^{M} w_j \mathcal{N}\left(\mathbf{o}_t \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\right)} \qquad (12)$$

The expected mean $E_i\{\mathbf{O}\}$ corresponds to the M step of the EM algorithm and can be estimated as (see Eq. (13)):

$$E_i\{\mathbf{O}\} = \frac{1}{n_i} \sum_{t=1}^{l} p(i \mid \mathbf{o}_t)\mathbf{o}_t \qquad (13)$$

- Maximum Likelihood Linear Regression (MLLR): This adaptation strategy takes the new data and updates the UBM's mean parameters to maximize the likelihood of the adaptation data [18]. The adaptation is achieved by means of a transformation matrix A applied to every extended mean vectors of the UBM, to obtain the adapted model. The adapted mean vectors can be estimated as (see Eq. (14)):

$$\hat{\boldsymbol{\mu}}_i = \mathbf{A}\tilde{\boldsymbol{\mu}}_i \qquad (14)$$

where $\tilde{\boldsymbol{\mu}}_i$ is the extended mean vector $[1, \boldsymbol{\mu}_i]$, required in order to include the bias term during the linear regression. The $\rho \times (\rho + 1)$ matrix A is estimated using a EM algorithm with auxiliar function given by (see Eq. (15)) [18]:

$$\mathcal{Q}(\Theta, \Theta') = \mathcal{L}(\mathbf{O} \mid \Theta) \sum_{i=1}^{M} \sum_{t=1}^{l} p(i \mid \mathbf{o}_t)[k + \log |\boldsymbol{\Sigma}_i| + d_{ti}] \qquad (15)$$

where k is the constant $\rho \log(2\pi)$, and d_{ti} is the argument of the Gaussian function given by .

$$d_{ti} = \left(\mathbf{o}_t - \mathbf{A}\tilde{\boldsymbol{\mu}}_i\right)^T \boldsymbol{\Sigma}_i^{-1} \left(\mathbf{o}_t - \mathbf{A}\tilde{\boldsymbol{\mu}}_i\right)$$

GMM-Support Vector Machine

This method was proposed in [11] and combines the modelling properties of the GMM-UBM scheme, with the discrimination capabilities of the Support Vector Machines (SVM) [19]. The method consists on building an UBM in the same way as the former approach, but unlike the GMM-UBM where the adapted model is class-dependent (genuine/impostor), in this case one adaptation per each single signature is performed, and the mean vectors of the adapted GMM are used to construct a GMM supervector

that becomes in the new feature vector representing the signature, i.e. the GMM-UBM is used as a mapping between the original feature sequence $\mathbf{O}$ and the super vector $\mathbf{m} = [\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2, ..., \hat{\boldsymbol{\mu}}_M]$. Finally, the feature space constructed by stacking all the supervectors from the different signatures is used to feed a conventional classification stage based on SVM.

Variational Bayesian GMM

The Variational GMM (VGMM) is a hierarchical Bayesian model in which the parameters of the GMM are treated as random variables themselves with their corresponding prior distributions. The prior distribution imposed over the mixing coefficients $\mathbf{w} = [w_1, w_2, ..., w_M]$ is a Dirichlet distribution, where by symmetry, the same parameter α_0 is used for each of the components (see Eq. (16)). The parameter α_0 can be interpreted as the effective number of observations associated with each component in the mixture [14]. Similarly, the method introduces an independent Gaussian-Wishart prior governing the mean and precision (the inverse of the covariance matrix $\Lambda = \Sigma^{-1}$) of each Gaussian component (see Eq. (17)) [14]. Formally,

$$p(\mathbf{w}) = \mathrm{Dir}(\mathbf{w} \mid \alpha_0) = C(\alpha_0) \prod_{k=1}^{M} w_k^{\alpha_0 - 1} \qquad (16)$$

$$p(\boldsymbol{\mu}, \boldsymbol{\Lambda}) = \prod_{k=1}^{M} \mathcal{N}\left(\boldsymbol{\mu}_k \mid \mathbf{b}_0, (\beta_0 \boldsymbol{\Lambda}_k)^{-1}\right) \mathcal{W}\left(\boldsymbol{\Lambda}_k \mid \mathbf{W}_0, v_0\right) \qquad (17)$$

where $C(\alpha_0)$ is the normalisation constant for the Dirichlet distribution, and $\{\mathbf{b}_0, \beta_0, \mathbf{W}_0, v_0\}$ are the hyperparameters of the Gaussian-Wishart distribution. $\mathbf{b}_0$ is the hypermean of the mean distribution which is typically set to 0 by symmetry [14], β_0 is a scaling factor, $\mathbf{W}_0^{-1}$ is the scale matrix, and v_0 is called the "degrees of fredom", which must satisfy $v > \rho - 1$. v_0 controls how strong the confidence is on the prior $\mathbf{W}_0$ [20]. The training of this model can be achieved by an analogous algorithm to the EM called Variational EM (VEM), which, as the EM algorithm, also requires a proper initialisation of the hyperparamters [21]. The optimisation of the variational posterior distribution can also be split into two steps: the E step, where the current distributions over the model parameters are used to estimate the responsibility r_{tk} of the component k for generating the data point t, and the M step, where r_{tk} is used to re-estimate the parameters of the observed data (Θ), analogously to the conventional EM algorithm, and the new hyperparameters of the Dirichlet and Gaussian-Wishart distributions. The re-estimation formula's for the hyperparameters are given by (see Eqs. (18-22)) [14]:

$$\alpha_k = \alpha_0 + n_k \qquad (18)$$

$$\beta_k = \beta_0 + n_k \tag{19}$$

$$b_k = \frac{1}{\beta_k}(\beta_0 b_0 + n_k \mu_k) \tag{20}$$

$$W_k^{-1} = W_0^{-1} + n_k \Sigma_k + \frac{\beta_0 n_k}{\beta_0 + n_k}(\mu_k - b_0)(\mu_k - b_0)^T \tag{21}$$

$$v_k = v_0 + n_k \tag{22}$$

where $n_k = \sum_{t=1}^{l} r_{tk}$. For some observation **x,** the predictive distribution of this Bayessian model can be approximated as a mixture of Students t-distributions given by (see Eqs. (23) and (24)) [14]:

$$p(\mathbf{x}) = \sum_{k=1}^{M} \alpha_k \mathbf{St}(\mathbf{x} | \mathbf{b}_k, \mathbf{L}_k, v_k + 1 - \rho) \tag{23}$$

where the precision is given by

$$\mathbf{L}_k = \frac{(v_k + 1 - \rho)\beta_k}{(1 + \beta_k)} \mathbf{W}_k \tag{24}$$

According to [14], when the size of the data set is large, the predictive distribution Eq. (23) reduces to a mixture of Gaussians. Given an observation $\mathbf{o}_t$, the predictive distribution obtained from Eq. (23), can be used to estimate the responsibility of each component (similar to Eq. (12)) as (see Eq. (25)):

$$\hat{p}(k | \mathbf{o}_t) = \frac{\alpha_k \mathbf{St}(\mathbf{o}_t | \mathbf{b}_k, \mathbf{L}_k, v_k + 1 - \rho)}{\sum_{j=1}^{M} \alpha_j \mathbf{St}(\mathbf{o}_t | \mathbf{b}_j, \mathbf{L}_j, v_j + 1 - \rho)} \tag{25}$$

Using this responsibility and the re-estimation formula's for the k-th component of the M step in the VEM algorithm given by (see Eqs. (26) and (27)):

$$w_k = \frac{n_k}{l} = \frac{1}{l}\sum_{t=1}^{l} \hat{p}(k | \mathbf{o}_t) \quad \mu_k = \frac{1}{n_k}\sum_{t=1}^{l} \hat{p}(k | \mathbf{o}_t)\mathbf{o}_t \tag{26}$$

$$\Sigma_k = \frac{1}{n_k}\sum_{t=1}^{l} \hat{p}(k | \mathbf{o}_t)(\mathbf{o}_t - \mu_k)(\mathbf{o}_t - \mu_k)^T \tag{27}$$

it is possible to estimate the parameters of a conventional UBM. Furthermore, from it, any of the two former strategies and/or adaptations can be applied [22].

3. Experiments and results

3.1. Database

The data set used for the experiments was the MCYT100 Signature sub-corpus, which contains 25 client signatures and 25 highly skilled forgeries (with natural dynamics) from 100 signers. Both, on-line information (pen trajectory, pen pressure and pen azimuth=altitude), and off-line information (image of the written signature) are included in the database. Nevertheless, in this work only the on-line part of the database was used. By considering the total number of signers and signatures, the number of available samples for simulation are 100 x (25 + 25) = 5000 [23].

3.2. Experimental setup

All the experiments were performed using a bootstrapping validation methodology, with ten repetitions. The size of training and testing subsets were adjusted from 90%-10% to 20%-80% respectively, in order to evaluate the sensitivity of the methods to the number of training samples, and to simulate more real operation conditions. It is worth noting that there are 25 genuine signatures per client in the database, therefore, 90%-10% and 20%-80% corresponds to 22 - 3 and 5 - 20 signatures respectively. Samples used during training are not involved in testing.

Three different experiments were performed: in the first one, genuine and impostor signatures per user were used for training two different models (following each of the strategies exposed in Section 2.2); the decision was taken by estimating the likelihood ratio (Eq. (6)) and comparing it with respect to the Equal Error Rate (EER) decision threshold. This is an unrealistic scenario because it uses the impostor samples during training, but it is used here only for comparison purposes. In the next experiments, the genuine signatures from all the users were used to train a UBM model, from which genuine models per user were adapted. The validation was performed using genuine signatures from other signers (random forgeries - called experiment 2) and impostor samples (skilled forgeries - called experiment 3). The likelihood ratio in this case was estimated between the genuine model and the UBM, and also comparing it against the EER threshold. The number of Gaussians for the UBM model was evaluated in the range between 5 and 20. The kernel function for the SVM based model was a Radial Base Function (RBF) and the regularisation and kernel parameters were set during validation. The results are presented in terms of false acceptance rate (FAR), false rejection rate (FRR) and EER, along with the area under the Receiver Operator Characteristic curve (AUC) and Detection Error Trade-off (DET) plots.

3.3. Results

Table 1 shows the results for all the models and strategies described in Section 2.2, according to the experiment

one. It is possible to observe how the performance of the models degrades with respect to a reduction in the number of training samples. However, it is worth to highlight that the system based on GMM-UBMMAP method was able to keep the accuracy above 93%, even when only 20% of the samples were used for training. The accuracy obtained by the GMM-SVM models, is considerable better than the GMM-UBM models when the available training subset is at least 50% of the whole database (corresponding to 13 signatures). When 90% of the samples were used for training, the relative reduction of the recognition rate using the GMM-SVM model is of 35.26% (2.57% in absolute terms). On the contrary, the performance of GMM-SVM models degrades faster for the experiments with 40% or less training data. The best result is achieved by the combination of the variational learning of UBM, along with a MAP adaptation and a SVM supervector classifier. This scheme yields to a recognition rate above 99% when the training included more than 80% of the samples. However, similar to the GMM-SVM models, its performance degrades quite fast with 40% or less training data. This could be explained because even though the training of the UBM is carried out using a full Bayesian method, this scheme requires two different adaptations, which demand enough data. It is worth to note that in this context the MAP adaptation always provided better results than the MLLR.

Figure 1 shows the accuracy obtained by all the methods evaluated according to the experiment one. As it was pointed out above, the best result was achieved by the combination of a variational GMM with a SVM, through a MAP adaptation, which following the notation in [11], could be called a Variational GMM-Supervector. Nevertheless, the standard GMM-UBM model was the most stable with respect to the size of the training subset. Anyway, all the models based on a UBM were better than the conventional GMM for small training sets.
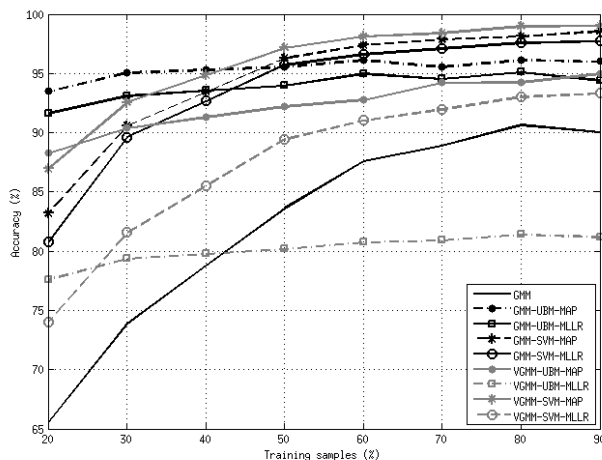

Figure 1 Accuracy obtained for all the methods evaluated according to the experiment one

Figures 2 and 3 show the DET curves for all the methods evaluated according to the experiment one, using 50% and 20% of the samples for training respectively.
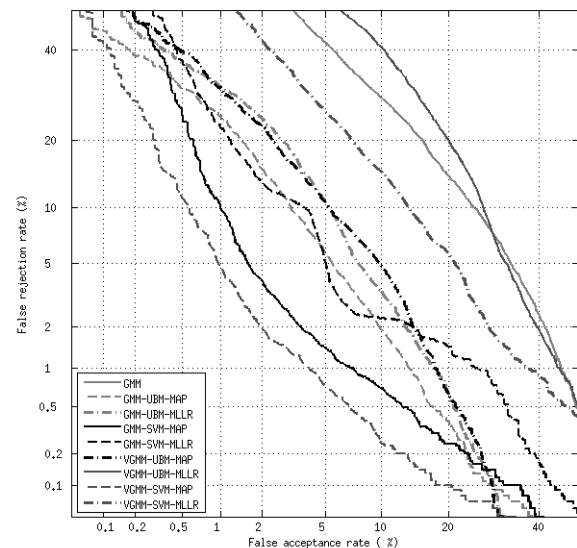

Figure 2 DET curves obtained for all the methods evaluated according to the experiment one, using 50% of the samples for training
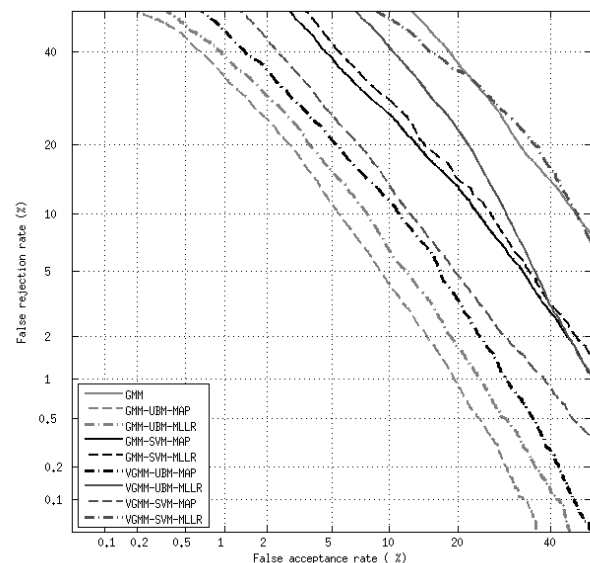
Figure 3 DET curves obtained for all the methods evaluated according to the experiment one, using 20% of the samples for training

Table 2 shows the results for all the models according to the experiment two. This is a more realistic scenario in which there is not available impostor samples during the training stage. It is of course, a more difficult challenge for the system. As it was pointed out before, during the experiment two the system is tested using the validation subset of one genuine signer (positive class) against the validation set of all other signers (negative class). This process is repeated for all the signers in the database. Therefore, there are much more samples in the negative class than in the

Table 1 FAR, FRR, EER, and AUC for all the methods evaluated according to the experiment one

Training size	System	Performance			AUC
		FAR	FRR	EER	
90%	GMM	6.20% ± 8.03*	13.7% ± 9.18	9.93% ± 7.57	0.93
	GMM-UBM-MAP	4.47% ± 7.08	3.47% ± 6.05	3.97% ± 5.00	0.99
	GMM-UBM-MLLR	6.10% ±8.71	5.03% ±8.26	5.57% ±6.40	0.99
	GMM-SVM-MAP	1.20% ±2.50	1.60% ±3.20	1.40% ±2.27	1.00
	GMM-SVM-MLLR	2.13% ±3.27	2.40% ±4.11	2.27% ±2.85	0.99
	VGMM-UBM-MAP	4.83% ±6.92	5.17% ±7.56	5.00% ±6.77	0.99
	VGMM-UBM-MLLR	19.3% ±15.3	18.2% ±13.9	18.77% ± 13.76	0.89
	VGMM-SVM-MAP	1.13% ±2.07	0.77% ±1.76	0.95% ± 1.65	1.00
	VGMM-SVM-MLLR	8.17% ±12.6	5.17% ±8.91	6.67% ±9.85	0.96
70%	GMM	6.23% ±6.72	16.0% ±9.17	11.09% ±7.21	0.92
	GMM-UBM-MAP	4.62% ±5.0	4.21% ±4.53	4.42% ±3.71	0.99
	GMM-UBM-MLLR	5.79% ±5.44	5.10% ±5.11	5.44% ±3.99	0.98
	GMM-SVM-MAP	1.80% ±2.45	2.30% ±3.14	2.10% ±2.39	1.00
	GMM-SVM-MLLR	2.72% ±3.44	3.06% ±4.02	2.89% ±3.10	0.99
	VGMM-UBM-MAP	5.95% ±7.02	5.61% ±6.80	5.78% ±6.64	0.99
	VGMM-UBM-MLLR	19.7% ±14.2	18.4% ±14.0	19.08% ± 13.71	0.88
	VGMM-SVM-MAP	1.86% ±2.40	1.29% ±2.15	1.57% ± 2.00	1.00
	VGMM-SVM-MLLR	10.6% ±11.1	5.49% ±9.60	8.02% ± 9,45	0.95
50%	GMM	8.49% ±7.94	24.2% ±8.87	16.35% ±7.35	0.89
	GMM-UBM-MAP	4.64% ±4.08	4.18% ±3.68	4.41% ±3.22	0.98
	GMM-UBM-MLLR	6.20% ±4.38	5.90% ±4.22	6.00% ±3.50	0.98
	GMM-SVM-MAP	3.38% ±3.44	4.04% ±3.88	3.71% ±3.30	0.99
	GMM-SVM-MLLR	4.42% ±4.46	4.09% ±4.11	4.26% ±3.69	0.97
	VGMM-UBM-MAP	7.87% ±7.38	7.68% ±8.03	7.77% ±7.42	0.98
	VGMM-UBM-MLLR	20.7% ±14.2	18.9% ±13.4	19.79% ± 13.25	0.88
	VGMM-SVM-MAP	2.90% ±3.56	2.73% ±3.38	2.82% ± 3.13	0.99
	VGMM-SVM-MLLR	14.00% ±13.9	7.09% ±11.6	10.53% ±11.84	0.93
30%	GMM	14.6% ±10.97	37.5% ±9.34	26.07% ±7.30	0.83
	GMM-UBM-MAP	4.96% ±3.43	4.88% ±3.58	4.92% ± 3.00	0.98
	GMM-UBM-MLLR	6.80% ±3.89	6.99% ±4.28	6.86% ±3.54	0.97
	GMM-SVM-MAP	8.05% ±6.74	10.8% ±7.42	9.42% ±6.35	0.96
	GMM-SVM-MLLR	9.45% ±7.99	11.3% ±8.31	10.39% ± 7.55	0.95
	VGMM-UBM-MAP	9.47% ±8.51	9.72% ±9.17	9.59% ±8.44	0.97
	VGMM-UBM-MLLR	20.2% ±13.0	21.0% ±14.4	20.61% ± 13.33	0.87
	VGMM-SVM-MAP	7.59% ±5.73	7.31% ±5.75	7.45% ±4.99	0.97
	VGMM-SVM-MLLR	20.6% ±16.0	16.3% ±16.8	18.41% ±14.33	0.85
20%	GMM	19.4% ±11.17	49.5% ±9.28	34.46% ±6.63	0.77
	GMM-UBM-MAP	6.40% ±3.93	6.63% ±4.21	6.51% ± 3.49	0.98
	GMM-UBM-MLLR	8.40% ±4.00	8.40% ±4.52	8.37% ±3.60	0.97
	GMM-SVM-MAP	14.8% ±10.5	19.0% ±11.0	16.79% ±10.03	0.90
	GMM-SVM-MLLR	15.6% ±11.3	22.9% ±12.5	19.22% ± 11.10	0.89
	VGMM-UBM-MAP	11.4% ±9.73	12.1% ±9.85	11.74% ± 9.38	0.95
	VGMM-UBM-MLLR	21.6% ±14.6	23.2% ±14.8	22.37% ± 13.89	0.86
	VGMM-SVM-MAP	12.4% ±8.73	13.7% ±8.76	13.03% ± 7.90	0.93
	VGMM-SVM-MLLR	26.2% ±17.7	25.7% ±19.8	25.96% ±16.02	0.78

* mean ± standard deviation

positive one. For instance, in the experiment 20%-80%, the model of every signer is tested using 20 genuine and 1980 impostor signatures. Bearing this in mind, it is possible to observe that several of the evaluated models were able to detect perfectly impostor signatures. Actually, the models based on GMM-SVM-MAP and VGMM-SVM-MAP, were able to detect all the impostor signatures even when only 20% of the samples were used for training. On the other hand, it is also possible to observe that as long as the number of training samples decreases, the system becomes more and more biased to the general class represented by the UBM. This fact is even more evident for systems using the combination of GMM and SVM models, whilst FAR remains in low rates, FRR increases to very high levels. This fact can be explained because during training, the samples used as positive class correspond to the genuine samples from one user, while the negative class is formed by the genuine samples from all other users in the database. This configuration produces an unbalanced training set that skews the system to the class with more data, reducing false positives and increasing false negatives. Although the percentage of training samples is reduced in the same proportion for both classes, this behaviour becomes more evident as the percentage of the training data is reduced because the number of genuine samples reaches critical values. Nevertheless, it is worth noting that the FRR obtained by VGMM-SVM-MAP during the experiment 20%-80% (38.2%) equates to 7.6 samples, i.e. the system made, on average, 7.6 mistakes every 2000 validations. Moreover, although the aim is to get a system with FAR and FRR as low as possible, in the context of biometric verification systems is most important do not give access to impostor people (i.e. to achieve a low FAR), than to commit some errors with genuine users (i.e. to achieve a low FRR), which would be asked to perform a new verification.

From Table 2, it is also possible to observe clearly, the superiority of the system based on the variational GMM-supervector. It is important to note that, for almost all the models, the performance degrades faster when the training set included less than 50% of the samples (around 12 signatures per user), which could be explained by the intrinsic variability among the signatures patterns from a single user, preventing the system to capture enough discriminative information from very small data sets. Anyway, the combination of a full Bayesian UBM and a SVM, was able to achieve more than 99.6% of accuracy using only 5 samples for training, and 99.8% when 12 signatures were used instead.

Figures 4 and 5 show the DET curves for the methods evaluated according to the experiment two, using 50% and 20% of the samples for training respectively. Only the models based on MAP adaptation were included, since in all the cases MAP beats MLLR. From these figures, it is possible to observe that for 50% of the training samples, the performance of the GMM-SVM-MAP method is quite similar to the variational version of the same scheme. However, for 20% of training samples, there is a greater difference between these two methods, confirming that the Bayesian approach is more suitable for small-sample conditions.
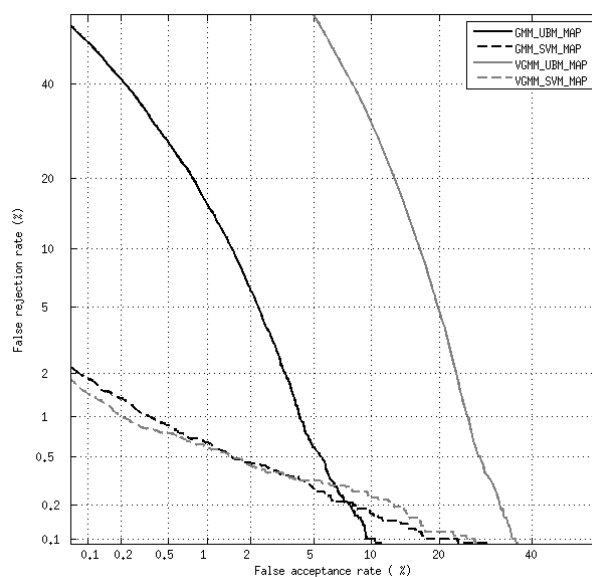

Figure 4 DET curves obtained for the methods based on MAP adaptation and evaluated according to the experiment two, using 50% of the samples for training
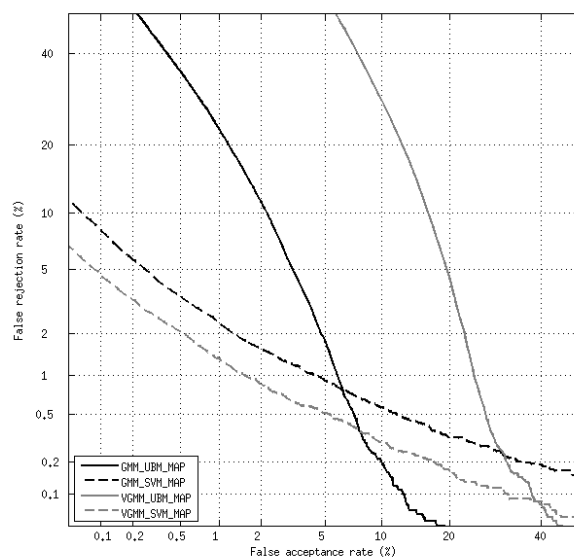

Figure 5 DET curves obtained for the methods based on MAP adaptation and evaluated according to the experiment two, using 20% of the samples for training

Table 3 shows the results for all the models according to the experiment three. This is the same system that in the experiment two but tested using the impostor samples per user. In this case, the imbalance is not as strong as in the former case, since in every repetition there are only 25 negative samples (the total number of impostor signatures per client in the database). Therefore, in this case the positive samples wrongly classified have more weight on the global error (we are aware that other performance measures

Table 2 FAR, FRR, EER and AUC for all the methods evaluated according to the experiment two (random forgeries)

Training size	System	Performance			AUC
		FAR	FRR	EER	
90%	GMM-UBM-MAP	1.06% ± 1.27	10.97% ±8.48	1.15% ±1.27	0.99
	GMM-UBM-MLLR	1.13% ± 1.38	9.5% ±6.42	1.22% ±1.37	0.99
	GMM-SVM-MAP	0% ± 0.00	12.03% ±6.90	0.12% ±0.07	1.00
	GMM-SVM-MLLR	0% ± 0.00	48.13% ±6.47	0.48% ±0.06	0.60
	VGMM-UBM-MAP	0% ± 0.00	11.57% ±8.48	0.11% ±0.08	1.00
	VGMM-UBM-MLLR	34.57% ± 16.79	35.9% ±17.87	34.58% ±16.79	0.49
	VGMM-SVM-MAP	0% ± 0.00	11.1% ±7.9	0.11% ±0.08	1.00
	VGMM-SVM-MLLR	0.4% ± 0.29	29.9% ±27.01	0.69% ±0.44	0.99
70%	GMM-UBM-MAP	0.41% ±0.59	12.62% ±7.67	0.53% ±0.60	1.00
	GMM-UBM-MLLR	1.53% ±1.56	11.97% ±6.86	1.64% ±1.54	0.99
	GMM-SVM-MAP	0% ±0.00	7.58% ±9.15	0.18% ±0.07	1.00
	GMM-SVM-MLLR	2.18% ±1.12	49.96% ±5.15	2.65% ±1.10	0.60
	VGMM-UBM-MAP	10.98% ±7.12	11.1% ±6.92	10.99% ±7.08	0.93
	VGMM-UBM-MLLR	37.78% ±16.77	38.36% ±16.74	37.793% ±16.76	0.58
	VGMM-SVM-MAP	0% ±0.00	12.69% ±6.59	0.12% ±0.07	1.00
	VGMM-SVM-MLLR	0.38% ±0.36	33.41% ±27.04	0.71% ±0.48	0.99
50%	GMM-UBM-MAP	0.71% ±0.84	13.55% ±5.75	0.83% ±0.85	1.00
	GMM-UBM-MLLR	2.14% ±1.79	12.92% ±6.08	2.24% ±1.79	0.99
	GMM-SVM-MAP	0% ±0.00	22.78% ±8.96	0.23% ±0.09	1.00
	GMM-SVM-MLLR	2.68% ±0.62	27.7% ±8.22	2.93% ±0.61	0.90
	VGMM-UBM-MAP	11.15% ±7.30	13.18% ±7.03	11.17% ±7.26	0.92
	VGMM-UBM-MLLR	35.99% ±16.67	38.16% ±15.69	36.01% ±16.65	0.60
	VGMM-SVM-MAP	0% ± 0.00	19.02% ±8.59	0.19% ±0.08	1.00
	VGMM-SVM-MLLR	0.3% ±0.25	41.61% ±29.37	0.70% ±0.41	0.99
30%	GMM-UBM-MAP	0.71% ±0.71	21.88% ±7.66	0.93% ±0.71	0.99
	GMM-UBM-MLLR	1.7% ±1.45	21.49% ±6.24	1.90% ±1.44	0.99
	GMM-SVM-MAP	0% ±0.00	33.58% ±10.69	0.34% ±0.10	1.00
	GMM-SVM-MLLR	4.09% ±0.99	36.74% ±9.73	4.41% ±1.00	0.90
	VGMM-UBM-MAP	8.29% ±5.59	20.39% ±7.26	8.41% ±5.56	0.93
	VGMM-UBM-MLLR	63.03% ±16.97	40.53% ±15.23	37.01% ±16.94	0.61
	VGMM-SVM-MAP	0% ± 0.00	29.27% ±11.15	0.29% ±0.11	1.00
	VGMM-SVM-MLLR	0.15% ±0.12	61.26% ±30.75	0.75% ±0.36	0.98
20%	GMM-UBM-MAP	1.03% ±1.15	27.93% ±8.04	1.3% ±1.15	0.99
	GMM-UBM-MLLR	2.37% ±1.88	29.76% ±9.17	2.64% ±1.87	0.98
	GMM-SVM-MAP	0% ±0.00	46.6% ±12.5	0.47% ±0.13	1.00
	GMM-SVM-MLLR	0.01% ±0.00	47.89% ±10.88	0.48% ±0.11	0.90
	VGMM-UBM-MAP	8.66% ±6.04	26.6% ±7.28	8.84% ±6.01	0.93
	VGMM-UBM-MLLR	37.20% ±17.26	40.6% ±17.28	37.23% ±17.22	0.56
	VGMM-SVM-MAP	0% ±0.00	38.2% ±12.33	0.38% ±0.12	1.00
	VGMM-SVM-MLLR	0.09% ±0.08	70.99% ±27.81	0.79% ±0.31	0.97

Table 3 FAR, FRR, EER and AUC for all the methods evaluated according to the experiment three (skilled forgeries)

Training size	System	Performance			AUC
		FAR	FRR	EER	
90%	GMM-UBM-MAP	14.28% ± 11.71	10.97% ±8.48	13.92% ± 10.66	0.87
	GMM-UBM-MLLR	16.27% ± 12.93	9.5% ±6.42	15.55% ±11.72	0.88
	GMM-SVM-MAP	2.87% ± 6.12	12.03% ±6.90	3.85% ±5.54	0.98
	GMM-SVM-MLLR	0.93% ± 2.48	48.13% ±6.47	5.99% ±2.48	0.60
	VGMM-UBM-MAP	21.34% ± 17.76	11.57% ±8.48	20.29% ± 16.03	0.84
	VGMM-UBM-MLLR	62.71% ± 17.70	35.9% ±17.87	59.84% ± 16.84	0.48
	VGMM-SVM-MAP	3.32% ± 6.73	11.1% ±7.9	4.15% ±6.13	0.98
	VGMM-SVM-MLLR	14.73% ± 16.18	29.9% ±27.01	16.36% ±14.25	0.88
70%	GMM-UBM-MAP	12.78% ± 12.68	12.62% ±7.67	12.75% ±9.91	0.86
	GMM-UBM-MLLR	15.34% ± 12.94	11.97% ±6.86	14.53% ± 10.22	0.86
	GMM-SVM-MAP	1.94% ± 3.92	7.58% ±9.15	5.75% ±3.69	0.98
	GMM-SVM-MLLR	0.5% ± 1.4	49.96% ±5.15	12.49% ±1.67	0.60
	VGMM-UBM-MAP	20.99% ± 19.26	11.1% ±6.92	18.59% ±14.98	0.84
	VGMM-UBM-MLLR	66.57% ± 17.06	38.36% ±16.74	59.73% ± 15.80	0.47
	VGMM-SVM-MAP	2.72% ± 6.28	12.69% ±6.59	5.14% ± 5.32	0.98
	VGMM-SVM-MLLR	13.88% ± 15.96	33.41% ±27.04	18.62% ±12.34	0.87
50%	GMM-UBM-MAP	15.94% ± 11.81	13.55% ±5.75	15.13% ±8.48	0.86
	GMM-UBM-MLLR	14.7% ± 10.87	12.92% ± 6.08	14.09% ± 7.96	0.86
	GMM-SVM-MAP	3.4% ± 7.13	22.78% ± 8.96	10.03% ± 5.96	0.97
	GMM-SVM-MLLR	1.86% ± 3.85	27.7% ± 8.22	10.70% ± 4.07	0.88
	VGMM-UBM-MAP	22.3% ± 19.6	13.18% ± 7.03	19.18% ± 13.59	0.84
	VGMM-UBM-MLLR	54.17% ± 18.94	38.16% ±15.69	48.69% ± 15.77	0.50
	VGMM-SVM-MAP	2.6% ± 5.74	19.02% ±8.59	8.22% ±5.03	0.98
	VGMM-SVM-MLLR	11.92% ± 13.8	41.61% ±29.37	22.08% ±11.87	0.87
30%	GMM-UBM-MAP	10.06% ±9.75	21.88% ±7.66	15.01% ±6.69	0.88
	GMM-UBM-MLLR	12.07% ±10.88	21.49% ±6.24	16.01% ±6.75	0.86
	GMM-SVM-MAP	3.18% ±6.98	33.58% ±10.69	15.90% ±6.67	0.96
	GMM-SVM-MLLR	1.47% ±2.85	36.74% ±9.73	16.23% ±4.50	0.88
	VGMM-UBM-MAP	18.71% ±17.19	20.39% ±7.26	19.42% ± 10.97	0.83
	VGMM-UBM-MLLR	61.27% ±16.37	40.53% ±15.23	52.59% ± 13.85	0.48
	VGMM-SVM-MAP	1.17% ±3.63	29.27% ±11.15	13.26% ± 5.24	0.97
	VGMM-SVM-MLLR	7.53% ±10.04	61.26% ±30.75	30.02% ±12.40	0.84
20%	GMM-UBM-MAP	17.15% ±13.33	27.93% ±8.04	21.94% ±7.74	0.81
	GMM-UBM-MLLR	9.36% ± 9.08	29.76% ± 9.17	18.43% ± 6.41	0.87
	GMM-SVM-MAP	1.96% ± 4.11	46.6% ± 12.5	21.80% ± 6.34	0.96
	GMM-SVM-MLLR	1.02% ± 2.15	47.89% ± 10.88	21.85% ± 5.02	0.88
	VGMM-UBM-MAP	17.94% ± 16.32	26.6% ± 7.28	21.79% ± 10.30	0.82
	VGMM-UBM-MLLR	63.31% ±17.28	40.6% ±17.28	53.22% ± 14.44	0.48
	VGMM-SVM-MAP	1.17% ±2.74	38.2% ±12.33	17.63% ± 5.76	0.97
	VGMM-SVM-MLLR	5.8% ±8.86	70.99% ±27.81	34.77% ±11.78	0.83

such as the geometric mean or the probability excess are less sensitive to the relative class frequency in the test set. However, those kind of measures are not commonly used in the context of signature verification, so they cannot be used for comparison purposes. Anyway, since those kind of measures are usually estimated from sensitivity and specificity measures, they can be easily estimated from the FAR and FRR values provided in the tables if needed). The best performance was obtained by a system based on a GMM-UBM model with MAP adaptation, achieving on average, an error of 3.85% when the 90% of the samples were used for training. During the experiment 20%-80%, the best performance was obtained by the variational GMM-supervector. In this case, the FAR increases up to 1.17% in comparison to the 0.0% achieved in the former experiment. This means that on average, for every 100 validations with impostors signatures, the system accepted as genuine one. On the other hand, in this experiment the positive samples are the same than in the previous one, so the FRR values are exactly the same than in Table 2.

Figures 6 and 7 show the DET curves for the methods evaluated according to the experiment three, using 50% and 20% of the samples for training respectively. Once again, only the models based on MAP adaptation were included. Unlike the previous experiment, in this case the performance of the GMM-SVM-MAP and VGMM-SVM-MAP remain similar even during the experiment 20%-80%, with a slightly improvement of the variational method when the training samples were reduced up to 5 (20%).
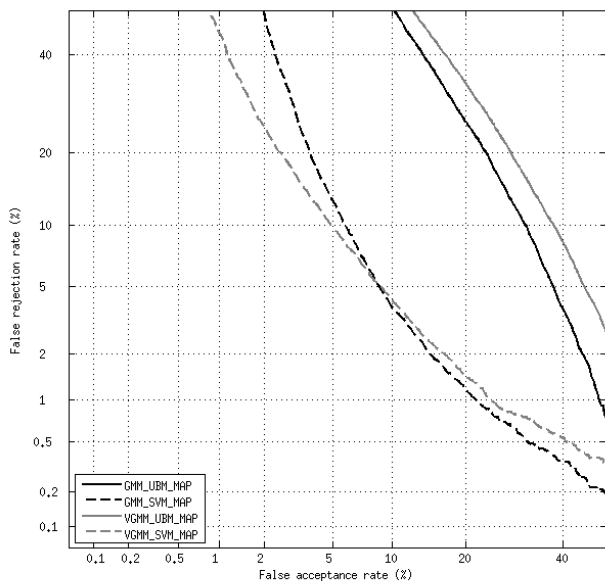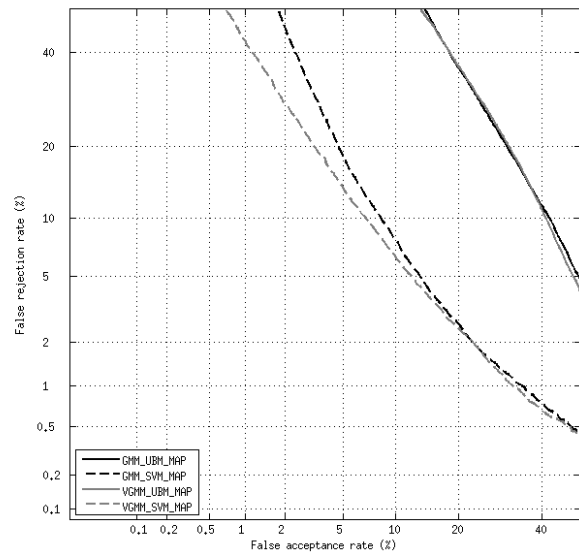

Figure 7 DET curves obtained for the methods based on MAP adaptation and evaluated according to the experiment three, using 20% of the samples for training

3.4. Discussion

A comparison of the performance of different signature verification systems is a difficult task since each author constructs his own signature data-sets. The lack of a standard international signature database continues to be a major problem for performance comparison. For the sake of completeness, in Table 4 we present some results obtained by published studies that used the MCYT database. Although it is not possible to carry out a direct comparison of the results, since the methodologies of training and testing and the classification strategies used by each author are different, Table 4 enables one to visualise results from the proposed methodology along side results published by other authors. It is worth to highlight that, most of the papers include some percentage of impostor samples into the training set, whilst in our experiments, assuming that in more realistic conditions there are not impostor samples available, impostor samples were only used during testing. Results presented here could be considered acceptable, taking into consideration that only raw data (i.e. velocity, acceleration) were used to feed the classifier, while most of the papers use more advanced characterisation strategies. In the present work, the main aim is to test the generalization capabilities of the Bayesian learning techniques, presented in section 2.2, in the context of on-line signature verification; nevertheless, the next step is to combine this kind of learning techniques with more robust characterisation strategies, for which more direct comparisons can be made.

Figure 6 DET curves obtained for the methods based on MAP adaptation and evaluated according to the experiment three, using 50% of the samples for training

Table 4 Comparison

	# sample per training	%EER
[10]	5	8.11
[24]	5	4.5
[25]	5	4.02
[26]	5	5.29
[27]	5	5.37
Proposed (GMM-UBM-MAP)	5	6.51

4. Conclusion

The paper evaluates nine different strategies based on Gaussian Mixture Models in the on-line signature verification task. All the models were tested under different conditions of available training samples and three different experiments including skilled forgeries during training and testing, and only in testing.

For almost all the methods evaluated, the performance degrades faster when the training set included less than 50% of the samples (around 12 signatures per user), which can be explained by the intrinsic variability among the signature patterns from a single user, preventing the system to capture enough discriminative information from very small data sets. However, for the genuine vs. impostor experiment, the GMM-UBM model was able to keep the equal error rate around 6% even when only 5 signatures per user were used in the training set. However, efforts to enhance feature extraction should be made.

In almost all the cases, the VGMM-UBM-SVM, was the model with the best performance, confirming that the Bayesian learning is more suitable for small-sample size conditions. This model was able to keep the false acceptation rate lower than 3% using only 5 signatures per used for training, and without any information about skilled forgeries. Moreover, when the system was tested against an impostor claiming to be another user but tracing his own signature, the system based on the combination of GMM-SVM, either using EM or Variational EM, achieved FARs equal to 0%, even for 20% of training samples, i.e. the systems rejected all the impostors without fail.

For all the experiments, the performance of MAP adaptation was by far better than the MLLR one. The combination of the classification strategies based on GMM-SVM and VGMM-SVM, with more advanced characterisation methods, should be the next step to figure out the real potential of these methods in the on-line signature verification task.

5. Acknowledgment

6. References

1. S. Liu and M. Silverman, "A practical guide to biometric security technology", *IT Professional*, vol. 3, no. 1, pp. 27-32, 2001.
2. K. Franke and J. Ruiz, "Soft-biometrics: Soft-computing technologies for biometric-applications", in *AFSS International Conference on Fuzzy Systems*, Calcutta, India, 2002, pp. 171-177.
3. S. Impedovo and G. Pirlo, "Verification of handwritten signatures: An overview", in *14th International Conference on Image Analysis and Processing* (ICIAP), Modena, Italy, 2007, pp. 191-196.
4. J. Vargas, M. Ferrer, C. Travieso and J. Alonso, "Off-line signature verification based on grey level information using texture features", *Pattern Recognition,* vol. 44, no. 2, pp. 375-385, 2011.
5. M. Malekar and S. Patel, "Off-line signature verification using artificial neural network", *International Journal of Emerging Technology and Advanced Engineering,* vol. 3, no. 9, pp. 127-130, 2013.
6. M. Kumar, "Signature verification using neural network", *International Journal on Computer Science and Engineering,* vol. 4, no. 9, pp. 1498-1504, 2012.
7. E. Argones and J. Alba, "Online signature verification based on generative models", *IEEE Transactions on Systems, Man, and Cybernetics, Part B,* vol. 42, no. 4, pp. 1231-1242, 2012.
8. D. Reynolds, T. Quatieri and R. Dunn, "Speaker verification using adapted Gaussian Mixture Models", *Digital Signal Processing*, vol. 10, no. 1-3, pp. 19-41, 2000.
9. L. Wan and B. Wan, "On-line signature verification with two-stage statistical models", in *8th International Conference on Document Analysis and Recognition* (ICDAR), Seoul, South Korea, 2005, pp. 282-286.
10. M. Martinez, J. Fierrez and J. Ortega, "Universal background models for dynamic signature verification", in *1st IEEE International Conference on Biometrics: Theory, Applications, and Systems* (BTAS), Crystal City, USA, 2007, pp. 1-6.
11. W. Campbell, D. Sturim and D. Reynolds, "Support vector machines using GMM supervectors for speaker verification", *IEEE Signal Processing Letters*, vol. 13, no. 5, pp. 308-311, 2006.
12. H. Attias, "Inferring parameters and structure of latent variable models by variational Bayes", in *15th Conference on Uncertainty in Artificial Intelligence,* Stockholm, Sweden, 1999, pp. 21-30.
13. S. García, J. Luengo and F. Herrera, *Data Preprocessing in Data Mining, 1st* ed. New York, USA: Springer, 2015.
14. C. Bishop, *Pattern Recognition and Machine Learning, 1st* ed. New York, USA: Springer, 2006.
15. D. Reynolds, "Speaker identification and verification using Gaussian mixture speaker models", *Speech Communication*, vol. 17, no. 1-2, pp. 91-108, 1995.
16. R. Duda, P. Hart and D. Stork, *Pattern classification, 2nd* ed. New Jersey, USA: Wiley-Interscience, 2000.
17. M. Ferras, L. Cheung, C. Barras and J. Gauvain, "Comparison of speaker adaptation methods as feature extraction for SVM-based speaker recognition", *IEEE Transactions on Audio, Speech, and Language*

Processing, vol. 18, no. 6, pp. 1366-1378, 2010.

18. C. Leggetter and P. Woodland, "Maximum likelihood liner regression for speaker adaptation of continuous density hidden Markov models", *Computer Speech and Language*, vol. 9, no. 2, pp. 171-185, 1995.

19. C. Cortes and V. Vapnik, "Support-vector networks", *Machine Learning,* vol. 20, no. 3, pp. 273-297, 1995.

20. K. Murphy, *Machine Learning: A Probabilistic Perspective, 1st* ed. Cambridge, USA: MIT Press, 2012.

21. N. Nasios and A. Bors, "Variational learning for Gaussian Mixture Models", *IEEE Trans. Systems, Man, Cybern., Part B*, vol. 36, no. 4, pp. 849-862, 2006.

22. V. Sahu, H. Mishra and C. Shekar, "Variational bayes adapted GMM based for audio clip classification models", in *3rd Int. Conf. Pattern Recognition Mach. Intell.*, New Delhi, India, 2009, pp. 513-518.

23. J. Fierrez, J. Ortega, D. Torre and J. Gonzalez, "Biosec baseline corpus: A multimodal biometric database", *Pattern Recognition,* vol. 40, no. 4, pp. 1389-1392, 2007.

24. J. Montalvao, N. Houmani and B. Dorizzi, "Comparing GMM and parzen in automatic signature recognition a step backward or forward", in *XVIII Brazilian Congress on Automatics,* Bonito, Brazil, 2010, pp. 4463-4468.

25. N. Sae and N. Memon, "Online signature verification on mobile devices", *IEEE Transactions on Information Forensics and Security,* vol. 9, no. 6, pp. 933-947, 2014.

26. J. Fierrez, L. Nanni, J. López, J. Ortega and D. Maltoni, "An on-line signature verification system based on fusion of local and global information", in *5th International Conference on Audio- and Video-Based Biometric Person Authentication* (AVBPA), Hilton Rye Town, NY, USA, 2005, pp. 523-532.

27. S. Garcia *et al.*, "Online Handwritten Signature Verification", in *Guide to Biometric Reference Systems and Performance Evaluation, 1st* ed. D. Petrovska, G. Chollet and B. Dorizzi (eds). New York, USA: Springer, 2008, pp. 125-165.