# An improved model and feature set for signature recognition

Monica Carfagni, Matteo Nunziati
Dipartimento di Meccanica e Tecnologie Industriali
University of Florence
Via di Santa Marta, 3, Florence
ITALY
{monica.carfagni,matteo.nunziati}@unifi.it
http://www.dmti.unifi.it/CMpro-v-p-57.html

*Abstract:* - Automatic on-line signature recognition has been investigated by several authors in order to allow machines to recognize an user from its own biometric traits. The following paper deals with features and models required in order to allow a machine to learn and discriminate signatures. The proposed solution approaches the signature making process as the motion of a point in a bi-dimensional space and models statistic properties of the motion via the well known Maximum a Posteriori training of Gaussian Mixture Models. Comparing our approach to state-of-the-art solutions, major advancements have been found. As first, both system accuracy in signature discrimination and system resistance to forgeries have been double. Eventually, the proposed modeling technique leads to smaller templates, whose size is halved with respect to state-of-the-art alternatives.

*Key-Words*: on-line signature recognition, pattern recognition, biometry, maximum a posteriori

## 1 Introduction

Automatic signature recognition has been investigated by several authors (as instance [1], [2]) in order to allow machines to recognize an user from its own biometric traits. The usefulness of this approach can be envisaged, as instance, in e-commerce and remote transaction authorization. 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 [3], 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.

Signature recognition methods can be split in two main fields: on-line and off-line. 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.

## 2 Problem Formulation and state-of-the-art Models

State-of-the-art on-line recognition techniques involve a two stage procedure for a system to be able to identify 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). 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 a new signature is requested, features are extracted from this new data stream and compared with the stored model.

It is common to employ log-likelihood ratios in order to accept or reject a claimed identity as true or false (e.g. an impostor). Let $F = \{f_i \, with \;\; i = 1, \ldots, n\}$ be the feature set, with $i$ indicating the $i$-th sample acquired by a device at a fixed sampling frequency (usually less than 100 Hz), and $\Theta_0$ be a template, a similarity score $S_0$ is defined as:

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

where $P$ is the probability operator[1]. If an alternative template $\Theta_1$ exists, it is possible to estimate a second score $S_1$ and retrieve the log-likelihood ratio $LLR$ as:

$$LLR = \log\left(\frac{S_0}{S_1}\right) \tag{2}$$

Fixed an acceptance threshold $\theta$, if $LLR \geq \theta$ a subject is considered as the target of $\Theta_0$, that is he/she is the subject the template has been derived from. Otherwise, the subject is considered as the target of $\Theta_1$.
It is worth the trouble to provide a deeper explanation for both $\Theta_1$ and $\theta$. Usually $\Theta_1$ 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 $\Theta_0$ 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 $\theta$ 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 $\theta_{EER}$. 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 $\theta_{EER}$ is the value for which an Equal Error Rate (EER) is attained, that is FA equates FR.
In addition to FA and FR, test sessions are employed in order to estimate system accuracy via so called DET plots [4]. Such plots represent the different values of the (FA, FR) pair on a Cartesian plot, in order to describe the system behavior independently of a predefined threshold. Some of these plots will appear in this paper as proof of our hypotheses.

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 $N(x,\mu_i,\Sigma_i)$, a GMM based template is defined as:

$$\Theta = \sum_{i=j}^{k} \alpha_j N(x, \mu_j, \Sigma_j) \;\; with \;\; \alpha_j \in \mathbb{R}, \forall j \;\; constrained \; to \;\; \sum_{j=1}^{k} \alpha_j = 1 \tag{3}$$

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

1 $S$ is actually computed evaluating the probability density of the $i$-th feature given the template $\Theta_0$.

ISBN: 978-1-61804-000-8

$$P(f_i|\Theta)=\sum_{j=1}^{k}\alpha_j N(f_i,\mu_j,\Sigma_j) \tag{4}$$

The generation of a GMM template involves the computation of the unbiased estimators for each mean $\mu_j$ and covariance matrix $\Sigma_j$ (which is usually constrained to be diagonal) as well as the weights $\alpha_j$. This estimation has not closed form solution, therefore the well known iterative Expectation-Maximization (EM) algorithm [5] is employed for the task.

## 3 State-of-the-art Processing and Features

On-line signature recognition 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 ($\gamma$,$\varphi$). 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:

$$x_0=\frac{1}{n}\sum_{i=1}^{n}x_i \quad y_0=\frac{1}{n}\sum_{i=1}^{n}y_i \tag{5}$$

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

$$\beta=\frac{1}{n}\sum_{i=1}^{n}\arctan \dot{y}_i/\dot{x}_i \tag{6}$$

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

$$\dot{q}_i=\frac{1}{2}\cdot\sum_{\tau=1}^{2}\tau\cdot(q_{i+\tau}-q_{i-\tau})/\sum_{\tau=1}^{2}\tau^2 \tag{7}$$

where $q$ is a generic variable. Eventually, the whole data set is rotated and translated so that the new $\beta$ 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 $\delta$, the instantaneous velocities $v$. Previous work on this topic [1] has demonstrated the scarce usefulness of the pair ($\gamma$,$\varphi$), pointing out how the remaining variables can better encoding subject specific traits. The resultant feature vector is thus $w$ = [x, y, p, $\delta$, $v$] , where $\delta$ and $v$ are computed as:

$$\delta=\arctan \dot{y}_i/\dot{x}_i \quad v=\sqrt{\dot{x}^2+\dot{y}^2} \tag{8}$$

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 [6]. Therefore, the final 1x15 feature vector is $f$ = [$w$, $\Delta w$, $\Delta\Delta w$] = [$x$, $y$, $p$, $\delta$, $v$, $\dot{x}$, $\dot{y}$, $\dot{p}$, $\dot{\delta}$, $\dot{v}$, $\ddot{x}$, $\ddot{y}$, $\ddot{p}$, $\ddot{\delta}$, $\ddot{v}$ ]. Eventually, features are

ISBN: 978-1-61804-000-8

normalized so that each component of $f$ is mapped to a canonical normal distribution with zero mean and unitary variance.

# 3 Proposed enhancements

We strongly believe that a feature set can be as representative of a physical phenomenon as the employed features can provide physically motivated quantities: moving from this consideration, we have reviewed the signature process from a physical perspective, which is compatible with the general theoretical framework worked out in [6]. As matter of fact, 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: ($\ddot{x}$, $\ddot{y}$) - ($\ddot{\cdot}$) being the second order derivative over time.

If we generalize 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: ($x$, $y$, $\delta$, $\dot{x}$, $\dot{y}$, $\dot{\delta}$), where the added parameters account for the instantaneous tangent angle and angular velocity. Moving from this model and by adding the pressure information, we propose the following reduced 1x10 feature vector:

$$f' = [x, y, \delta, p, v, \dot{\delta}, \dot{p}, \dot{v}, \ddot{\delta}, \ddot{p}] \tag{9}$$

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

Concerning the employed model, other fields of biometry make wide use of the so named UBM-GMM model. This model has been introduced in speaker recognition in [7] and represents a special case of the Maximum A Posteriori (MAP) estimator for HMM parameters, described in [8] and extended in [9]. 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 [7] have made their system less sensitive to this issue.

The procedure, detailed in [7] and [8], 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:

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

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 [7] and [9]; namely in [7] an a priori set of weights is employed, while in [9] a more advanced adaptive method is presented. As the authors are not aware of any other work in this sense applied to signature recognition, we propose here the first iteration of our research where the a priori version of MAP is employed.

# 4 Experimental Setup and Results

In order to test our hypotheses, the myIDea database [10] 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 a separate set *T*, employed to simulate different claimed identities. 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 *T*) and system sensitivity to forgeries (*T* vs. *F*) have been evaluated. Result are reported in terms of DET plots for the *T vs. T* test, while FA is used for the evaluation of the *T vs. F* test.
Being literature results based on different data sets, a baseline model has been built. The *baseline* templates are obtained by training GMM with $k$=1024 components against the feature vectors described in section 2; the size $k$ has been selected as optimal on the basis of a posteriori EER analysis, as in [6]. A second model *SYS1* is composed by the same type of templates but the employed feature set is the one defined by eq. (9). Eventually, a third model *SYS2* has been trained against the same feature set of *SYS1* but enrollment has been carried out by using the MAP algorithm. It is common to reduce the MAP procedure so that only means are adapted, that is the UBM and the templates share the same $\Sigma$ and $\alpha$. This procedure has shown to provide very good results (compare [7] 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 new feature set, while comparing *SYS2* with *SYS1* the effects of the MAP algorithm will be pointed out.

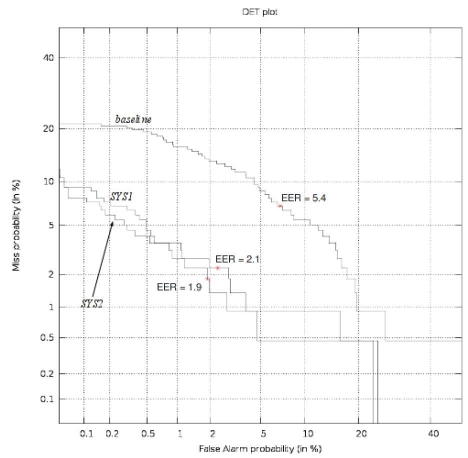


Fig.1: DET plot of the proposed systems (abscissas: FA, ordinates: FR).

As depicted in Fig.1, the employment of the novel feature set induces a dramatic improvement in system accuracy. As first observation, the EER is reduced from 5.4% to 2.1%, with a relative improvement of 61%, that is, the EER is halved. On the other side, the difference between *SYS1* and *SYS2* is neglectable and should be imputed to the statistic nature of the test. Indeed, templates are generated by using the EM (and MAP) algorithm, which assures the attainment of a local minimum and can generate different solutions on the basis of different initializations. It is common practice to initialize the EM with random starting points, therefore, different runs can generate slightly different results. Additionally, each system has been trained by shuffling signatures, that is, generating a random UBM data set and a random template feature set. As a consequence no major improvement appears in the experiment by using the MAP approach instead of a classical EM approach.

Different results can be observed in the *T vs. F* test, which assesses system's robustness to fake signatures. In this test, both the modified feature set and the MAP approach show relative improvements over the baseline system. Namely, the *baseline* system shows a FA of 11%; the new feature set induces a FA reduction of 27%

ISBN: 978-1-61804-000-8

(from 11% to 8%) with respect to *baseline*, while by adding the MAP approach we almost halved this error attaining a relative improvement of 45.5% (from 11% to 6%).
Considering also other effects of the MAP approach, additional benefits arise. If mean-only MAP is adopted in order to retrieve templates from an UBM, this implies that only templates means $\mu_i$ have to be stored on a device, as all other parameters are associated with an UBM, which commonly resides on a server. A well known side effect of this approach is that templates require a reduced amount of storage. Given a standard GMM model with diagonal covariance matrices, the size of a template is defined as: $k(2m+1)$ , where $k$ is the number of components and $m$ is the size of the feature vector. If only means have to be stored, only $km$ elements have to be retained in the client device.
This implies that, regardless of the optimal number of components, the ratio between the original model's size and the proposed model is $m_{new}/(2m_{old}+1)$ , where $m_{new}$ is 10 and $m_{old}$ is 15 in our case. This leads to a template size 51,6% smaller than in state-of-the-art solutions.

## 5 Conclusion

The following paper deals with features and models for on-line signature recognition. 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. Comparing our approach to state-of-the-art solutions, major advancements have been found. As first, both system accuracy in signature discrimination and system resistance to fake signatures (forgeries) have been double. Eventually, the proposed modeling technique leads to smaller templates, whose size is halved with respect to state-of-the-art alternatives.

*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, 2007, pp. 2325-2334.
2. J. Richiardi, A. Drygajlo, Gaussian Mixture Models for On-line Signature Verification, *Proceedings of the 2003 ACM SIGMM workshop on Biometrics*, 2003.
3. F.-M.E. Uzoka, T. Ndzinge, Empirical analysis of biometric technology adoption and acceptance in Botswana, *The Journal of Systems and Software,* No.82, 2009, pp. 1550–1564.
4. 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*, 1997, pp. 1895–1898.
5. 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, 1977, pp. 1–38.
6. 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*, 2009, pp. 1200–1208.
7. D.A. Reynolds, T.F. Quatieri, R.B. Dunn , Speaker Verification Using Adapted Gaussian Mixture Models, *Digital Signal Processing*, Vol.10, 2000, pp. 19–41.
8. 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, 1994, pp. 291–298.
9. P. Kenny, G. Boulianne, P. Dumouchel, Eigenvoice modeling with sparse training data, *IEEE Trans. Speech Audio Processing*, Vol. 13, No. 3, 2005 .
10. 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*, 2005, pp. 59-62.