# PERFORMANCE AND ROBUSTNESS: A TRADE-OFF IN DYNAMIC SIGNATURE VERIFICATION

*Javier Galbally, Julian Fierrez, and Javier Ortega-Garcia*

Biometric Recognition Group–ATVS, EPS, Universidad Autonoma de Madrid,
C/ Francisco Tomas y Valiente 11, 28049 Madrid, Spain
email: {javier.galbally, julian.fierrez, javier.ortega}@uam.es

## ABSTRACT

A performance and robustness study for on-line signature verification is presented. Experiments are carried out on the MCYT database comprising 16,500 signatures from 330 subjects, which are parameterized by means of a 100-feature set which can be divided into four different groups according to the signature information they contain, namely: $i$) time, $ii$) speed and acceleration, $iii$) direction, and $iv$) geometry. The SFFS feature selection algorithm is used to search for the best performing feature subsets under the skilled and random forgeries scenarios, and to find the most robust subsets against a hill-climbing attack. Comparative experiments are given, where it is shown that the most discriminant parameters are those regarding geometry information, while the most robust are the time related features.

***Index Terms***— *On-line* signature verification, attacks.

## 1. INTRODUCTION

On-line signature verification constitutes an intense research area due to its social and legal acceptance and the widespread use of the written signature as a personal authentication method [1]. Furthermore, it presents a high level of collectability, being easily acquired by means of different devices such as pen tablets, PDA's, Tablet PC's, etc. However, in spite of these advantages, on-line signature recognition still remains as a challenging problem specially due to its small inter-class variability (changes between the signatures produced by two different persons), and its high intra-class variations (changes between two signatures produced by the same user).

In order to solve the problems present in signature recognition, many efforts have been made to generate a compact set of features that maximizes the inter-class distance while minimizing the intra-class variations [2]. In addition to the mentioned intrinsic challenges present in the recognition of written signatures [3], on-line signature verification systems are also exposed to attacks which can decrease their level of security [4].

It would be desirable, when designing an on-line signature application, to have a set of features not only maximizing the inter-class while minimizing the intra-class variations, but also robust against these type of attacks.

With these premises it is clear that, in order to choose the best set of features possible for a particular signature recognition application, a trade-off between performance and robustness has to be reached. In the present contribution we analyze both aspects in an on-line signature verification system on the MCYT database [5], using the 100-feature set introduced in [2]. The SFFS feature selection algorithm [6] is used to search for the best performing feature subsets under the skilled and random forgeries scenarios, and to find the most robust subsets against the Bayesian hill-climbing attack described in [7]. Comparative experiments are given resulting in some findings on the most/least discriminant features for the scenarios considered, and the groups of features which are best suited to enhance/decrease the robustness of the system.

## 2. FEATURE EXTRACTION

The signatures are parameterized using the set of features described in [2]. In that work, a set of 100 global features was proposed. We have divided this set of parameters into four different groups according to the signature information they contain. All the features assigned to each class are specified in Table 1 (the numbering criterion is the same used in [2]).

One of the objectives of the present contribution is to give some indications on which of these parameter groups should be used to maximize the system performance and which are the best suited to increase the robustness of the application against the hill-climbing attack described in Sect. 4.2.

In the experimental study we analyze several subsets selected from the original 100-feature set. Due to the high dimensionality of the problem, exhaustive search is not feasible (there are $2^{100}$ possibilities to be explored). The feature selection method used in the experiments is the SFFS algorithm introduced in [6], which has shown remarkable performance

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

**Table 1**. Division of the feature set introduced in [2] according to the signature information they contain.

| | FEATURES |
|---|---|
| Time | 1,13,22,32,38,40–42,50,52,58-60,62,64,68,74,79,81–82,87-90,94,100. |
| Speed | 4–6,9–11,14,23,26,29,31,33,39,44–45,48,69,76,80,83,85,91-92,96. |
| Direction | 34,51,56-57,61,63,66,71–73,77-78,84,93,95,97–98,99. |
| Geometry | 2–3,7–8,12,15–21,24–25,27–28,30,35–37,43,46–47,49, 53–55,65,67,70,75,86 |

over other selection algorithms [8].

### 2.1. Signature Verification System

The signatures are parameterized using the set of features described in Sect. 2. In the present contribution we use this 100-feature representation of the signatures, normalizing each parameter to the range [0,1] using the tanh-estimators described in [9].

The similarity scores are computed as the inverse of the Mahalanobis distance between the input feature vector $\mathbf{y}$ and a statistical model of the client under consideration $\mathcal{C}$ (estimated using 5 training signatures).

## 3. HILL-CLIMBING ATTACK

In the present contribution we use the Bayesian approach to a hill-climbing attack presented in [7]. The core idea behind the algorithm is to iteratively adapt a known global distribution to the local specificities of the unknown user being attacked. For this purpose, a pool of signatures is used to compute the general statistical model $G$, which is sampled $N$ times. Each of the points in the distribution is compared with the client being attacked $\mathcal{C}$, generating $N$ similarity scores $J(\mathcal{C}, \mathbf{y}_i)$. The $M$ points which have generated higher scores are then used to compute a local distribution $L$, which is used to generate an adapted distribution $A$, that trades off (according to a parameter $\alpha$) the general knowledge provided by $G$ and the local information given by $L$. The global distribution is then redefined as $G = A$, and the process continues until the finishing criterion is met, i.e., one of the scores $J(\mathcal{C}, \mathbf{y}_i)$ exceeds the similarity threshold, or the maximum number of iterations is reached.

In Fig. 1 (b) we depict an example attack of the Bayesian hill-climbing algorithm using the best configuration reported in [7], i.e., $N = 50$, $M = 5$, and $\alpha = 0.4$. A lighter grey denotes the global distribution at iteration 1 (crosses), and 30 (triangles), respectively, while squares and circles are the points forming the local distributions (shown in a black thin line). The black thick ellipses are the adapted distributions at iterations 1 and 30, and the dashed line represents the attacked account. We can observe how the adapted distribution moves towards the target through the iterations (the attack was successful in iteration 60).

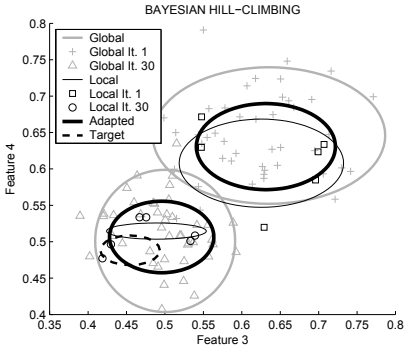



**Fig. 1**. Iterations 1 and 30 of an execution of the Bayesian hill-climbing algorithm described in [7].

## 4. DATABASE AND EXPERIMENTAL PROTOCOL

### 4.1. Data set Description

The experiments were carried out on the MCYT signature database [5], comprising 330 users. The database was acquired over 5 time-spaced capture sessions. Every client contributed with 25 genuine signatures and 25 skilled forgeries, to complete the 16,500 signatures that conform the database.

For each user, five different genuine models are computed using one training signature from each acquisition session.

### 4.2. Experimental Protocol

#### *4.2.1. Performance experiments*

The aim of these experiments is to find in the original 100-feature set, a number of subsets (each of a different dimension) which minimize the EER of the signature recognition system.

Two different scenarios are considered, $i$) *skilled forgeries*, in which the intruder tries to access the system imitating the original users's signature, and $ii$) *random forgeries*, where impostors try to access other's accounts using their own signature. In the first case, genuine scores are generated matching each of the five computed models of every user with the remaining 20 genuine signatures ($5 \times 20 \times 330 = 33,000$ genuine scores), while the impostor scores are computed comparing the 5 statistical models with all the 25 skilled forgeries, resulting in $5 \times 25 \times 330 = 41,250$ impostor scores. In the random forgeries scenario, genuine scores are computed as above, while each statistical model is matched with one signature of the remaining users to generate the $5 \times 330 \times 329 = 542,850$ impostor scores. These sets of genuine and impostor scores are then used to compute the EER of the system which is the criterion to be minimized in the SFFS algorithm.

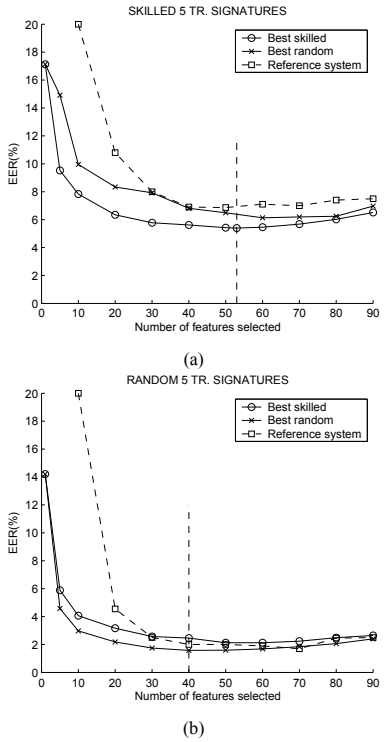



**Fig. 2**. System performance on the skilled (a), and random forgeries scenarios (b) using the SFFS feature subset selection maximizing the EER for skilled (circles), and random forgeries (crosses), compared to the reference system (squares) described in [2].

#### *4.2.2. Robustness experiments*

The objective of these experiments is to find a feature subset in the original 100 dimensional parameter space, which maximizes the robustness of the signature recognition system (i.e., minimizes the number of accounts bypassed) against the best configuration of the Bayesian hill-climbing algorithm described in [7].

In order to perform the robustness analysis, the database is divided into a training set (used to estimate the initial distribution $G$) and a test set comprising all the accounts being attacked, which are afterwards swapped (two-fold cross-validation). With this approach, a total $330 \times 5 = 1,650$ accounts are attacked. The number of broken accounts is used as the minimization criterion in the SFFS algorithm.

## 5. RESULTS

### 5.1. Performance Experiments

In Fig. 2, verification performance results for different subset sizes are given for the skilled forgeries scenario (a), and the random forgeries scenario (b). In circles we show the system performance when considering the subsets that perform best when coping with skilled forgeries (from now on, skilled subsets), while the system EER for the best random subsets is depicted with crosses. These results are compared to the on-line signature recognition system based on global features described in [2] (using a Parzen Windows based matcher and a top ranked selection scheme of best individual features).

As expected, the skilled subsets perform the best in the skilled forgeries scenario, while the random subsets minimize the EER in the random forgeries scenario. In both cases the combination of the Mahalanobis distance matcher and the SFFS feature selection outperforms the verification scheme described in [2], with relative improvements in the verification performance against skilled forgeries around 22% using 50 features, and more than 60% for small set sizes (10 features).

The curse of dimensionality is clearly patent in both figures, where the minimum EER has been highlighted with a vertical dashed line. The best performance point is reached for a 53 dimensional subset in the case of skilled forgeries (EER=5.39%), and for a subset comprising 40 features in the random forgeries scenario (EER=1.58%).

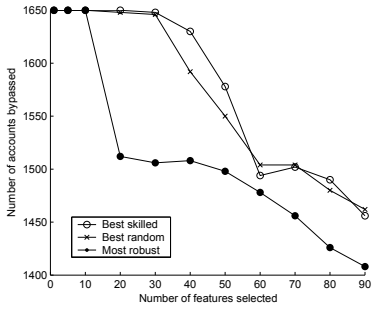



**Fig. 3**. Number of accounts bypassed for the skilled subsets (circles), the random subsets (crosses), and the feature subsets maximizing the robustness of the system (dots).

### 5.2. Robustness Experiments

In Fig. 3 we depict the number of accounts bypassed with the Bayesian hill-climbing attack described in [7] using the skilled (circles) and random subsets (crosses), and the most robust feature subsets found by the SFFS algorithm. Although the robust subsets show a better behaviour against the attack, none of the parameter sets show a significantly decrease in the system vulnerability, with only 15% of the accounts resisting the attack in the best case.

### 5.3. Comparative Experiments

The verification performance for the different subsets found in the previous experiments is shown in Fig. 4, both for the skilled (a), and the random forgeries scenarios (b). The circled solid line depicts the system EER for the skilled subsets,

**Table 2**. Number of features for the skilled, random, and robust subsets belonging to each of the groups described in Sect. 2.

| | Time | Speed | Direct. | Geomet. |
|---|---|---|---|---|
| Skilled | 2 | 2 | 0 | 1 |
| Random | 0 | 1 | 0 | 4 |
| Robust | 2 | 0 | 1 | 2 |

(a) Best 5-dimensional subsets.

| | Time | Speed | Direct. | Geomet. |
|---|---|---|---|---|
| Skilled | 3 | 3 | 0 | 4 |
| Random | 1 | 2 | 1 | 6 |
| Robust | 5 | 0 | 2 | 3 |

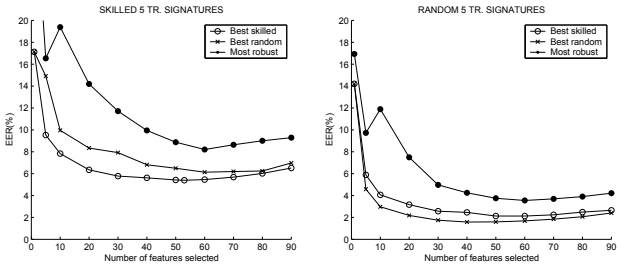
(b) Best 10-dimensional subsets.

| | Time | Speed | Direct. | Geomet. |
|---|---|---|---|---|
| Skilled | 6 | 5 | 7 | 12 |
| Random | 5 | 6 | 7 | 12 |
| Robust | 10 | 7 | 6 | 7 |

(c) Best 20-dimensional subsets.

the solid line with crosses represents the EER for the random subsets, while the dots indicate the system verification performance when using the robust subsets. It is clear from the results shown in both figures that the use of more robust sets of features leads to a significant decrease in the verification performance of the system.

In Table 2, we show the number of features belonging to each of the groups described in Sect. 2, for the different subsets (skilled, random and robust) found in the previous experiments. From this analysis we can see that the most robust features are those regarding time information while the most vulnerable are the speed related features. On the other hand, the most discriminant parameters are those containing geometry information, and the least discriminant the direction related features.




(a) (b)

**Fig. 4**. System performance on the skilled (a), and random scenarios (b) using the suboptimal subsets for skilled (circles) and random forgeries (crosses), and the subsets maximizing the system robustness (dots).

## 6. CONCLUSIONS

A performance and robustness study for an on-line signature verification system has been made. Experiments were carried out on the MCYT database (comprising 16,500 signatures) using the 100-feature set described in [2]. Several best performing feature subsets were found using the SFFS feature selection algorithm which outperform the global feature system described in [2] both for the skilled and random signatures scenarios. The SFFS algorithm was also used to search for parameter subsets which increase the robustness of the studied system against the hill-climbing attack described in [7]. It was shown experimentally that the most discriminant parameters are those containing geometry information, and the least discriminant the direction related features. On the other hand, the most robust features are those regarding time information while the most vulnerable are the speed related features.

Although a trade-off between performance and robustness should be reached, experiments show that the most robust subsets do not significantly decrease the system vulnerability compared to the best performing subsets, while the EER is clearly increased. Thus, it would be more advisable to search for parameter sets which improve the performance of the system, rather than those which enhance its robustness.

## 7. REFERENCES


[1] J. Fierrez and J. Ortega-Garcia, *Handbook of biometrics*, chapter On-line signature verification, Springer, to app.

[2] J. Fierrez-Aguilar, L. Nanni, et al., "An on-line signature verification system based on fusion of local and global information," in *Proc. of AVBPA*. 2005, LNCS-3546.

[3] H. Levi and V. Govindaraju, "A comparative study on the consistency of features in on-line signature verification," *Pattern Recognition Letters*, vol. 26, pp. 2483–2489, 2005.

[4] N.K. Ratha, J.H. Connell, and R.M. Bolle, "An analysis of minutiae matching strength," *Proc. AVBPA*, pp. 223–228, 2001.

[5] J. Ortega-Garcia, J. Fierrez-Aguilar, et al., "MCYT baseline corpus: a bimodal biometric database," *IEE VISP*, vol. 150, pp. 395–401, 2003.

[6] P. Pudil, J. Novovicova, and J. Kittler, "Flotating search methods in feature selection," *Pattern Recognition Letters*, pp. 1119–1125, 1994.

[7] J. Galbally, J. Fierrez, and J. Ortega-Garcia, "Bayesian hill-climbing attack and its application to signature verification," in *Proc. ICB*. 2007, pp. 386–395, LNCS-4642.

[8] A. K. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE Trans. PAMI*, vol. 19, pp. 153–158, 1997.

[9] A. K. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognition*, vol. 38, pp. 2270–2285, 2005.