Towards Mobile Authentication Using Dynamic Signature Verification: Useful Features and Performance Evaluation

Marcos Martinez-Diaz, Julian Fierrez, Javier Galbally, Javier Ortega-Garcia
Biometric Recognition Group - ATVS, EPS - Univ. Autonoma de Madrid
C/ Francisco Tomas y Valiente, 11 - Campus de Cantoblanco - 28049 Madrid, Spain
{*marcos.martinez, julian.fierrez, javier.galbally, javier.ortega*}*@uam.es*

Abstract

The proliferation of handheld devices such as PDAs and smartphones represents a new scenario for automatic signature verification. Traditionally, research on signature verification has been carried out employing signatures acquired using digitizing tablets or Tablet-PCs. In this paper we study the effects of the mobile acquisition conditions and we analyze the considerations that must be taken in the new handheld scenario. A signature verification system adapted to handheld devices via feature selection is proposed and a systematic comparison with a traditional pen tablet-based system is performed. The system is combined with another based on Hidden Markov Models using score fusion. Results confirm an increased signature variability in the case of handheld devices.[1]

1. Introduction

Despite its widespread social and legal acceptance, signature verification is still a challenging task within biometrics [1, 2]. As a behavioral biometric trait, signatures are subject to a considerable variability even on successive realizations, which can be increased over medium or large periods of time. Moreover, the possibility of creating forgeries with a relative ease, exposes a signature verification system to challenges not commonly present among other biometrics. Consequently, a signature verification system designer must face a high *intra-class* variability (between the signatures of a specific user) and a low *inter-class* variability, when forgeries are considered.

[1]This work has been supported by the Spanish Ministry of Education under project TEC2006-13141-C03-03. J. Fierrez is supported by a Marie Curie Fellowship from the European Commission. J. Galbally is supported by a FPU fellowship from the Spanish MEC.

Two main types of dynamic signature verification systems exist. *Feature-based* systems model the signature as a holistic multidimensional vector composed of global features [3]. *Function-based* systems extract time functions from the signature signal (pen coordinates, pressure, etc.) and perform signature matching via elastic or statistical techniques like Dynamic Time Warping (DTW) [4] or Hidden Markov Models (HMM) [5]. The typical architecture of an automatic signature verification system is depicted in Fig. 1.

Recently, smartphones and handheld devices have gathered a high level of popularity in the context of convergence and ubiquitous access to information and services. These devices represent a clear target for the deployment of a signature verification system, providing enough processing power and a stylus-based input. Signature verification can be used as a convenient alternative to passwords that may be forgotten or stolen for applications like e-commerce or access control. Nevertheless, signature verification on handheld devices is affected by factors not present in other input devices primarily due to a small input area, poor ergonomics or the fact that the user may be in movement. As a consequence, the signing process may be degraded.

Interestingly, the recent BioSecure Multimodal Evaluation Campaign (BMEC) [6], with the participation of independent research institutions, has shown that verification results for the case of handheld devices is significantly lower than those with other databases captured using a pen tablet [7].

In this paper, the problem of signature verification on handheld devices is studied. An analysis of the discriminant power of different types of features (temporal, geometric, etc.) is performed using the Fisher Discriminant Ratio (FDR) and feature selection algorithms. The resulting feature-based system, adapted to the handheld scenario, is further combined with an HMM system using score fusion, and the overall performance is measured against other sate-of-the-art systems using the re-

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

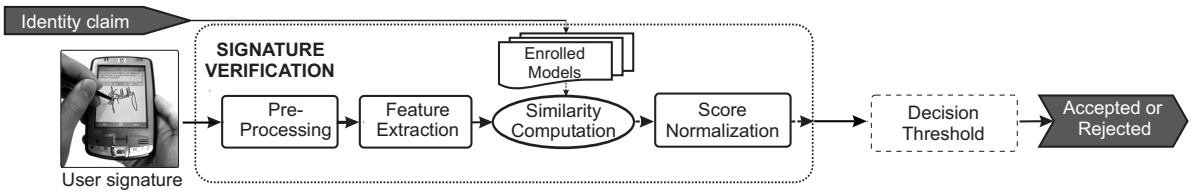

Figure 1. Signature Verification System Architecture.

sults of BMEC. The signatures used for experiments were captured both on a PDA and a digitizing tablet and correspond to the same users in both scenarios, allowing a fair comparison between them.

2. Related Work

Feature-based dynamic signature verification has been extensively studied [3, 8, 9]. Despite the large amount of different global feature sets that have been proposed (a maximum of 100 features are considered in [9]), the usually low amount of available training data motivates the use of feature selection to reduce the feature vector size (due to the curse of dimensionality). Several feature selection techniques have been proposed in the literature, being the Sequential Forward Feature Selection (SFFS) one of the best performing methods reported [10]. Nevertheless, to the extent of our knowledge, all the previous works on feature selection for dynamic signature verification have used data from digitizing tablets, so their observations may not fully apply to the case of handheld devices.

Function-based signature verification using DTW [4] or HMMs [5] is the most popular approach in signature verification. In these systems, the captured time functions are used to model each user signature. It must be taken into account that PDAs and other handheld devices are able to capture only pen position signals, while pen tablets provide additional signals such as pressure and pen inclination angles. An analysis of the implications of the lack of these signals in the PDA scenario is out of the scope of this work.

Finally, fusion of the feature- and function-based approaches has been reported to provide a better performance than the individual systems [9].

3. Signature Features

The set of features used in this work is the one presented in [9], which comprises 100 features. This is an extensive set that includes a considerable amount of features previously presented in the literature. These can be divided in four categories corresponding to the following magnitudes (the numbering is the same used in [9]):

- **Time** (25 features), related to signature duration, or timing of events such as pen-ups or local maxima: 1, 13, 22, 32, 38, 40-42, 50, 52, 58-60, 62, 64, 68, 79, 81-82, 87-90, 94, 100.
- **Speed and Acceleration** (25 features), from the first and second order time derivatives of the position time functions, like average speed or maximum speed: 4-6, 9-11, 14, 23, 26, 29, 31, 33, 39, 44-45, 48, 69, 74, 76, 80, 83, 85, 91-92, 96.
- **Direction** (18 features), extracted from the path trajectory like the starting direction or mean direction between pen-ups: 34, 51, 56-57, 61, 63, 66, 71-73, 77-78, 84, 93, 95, 97-99.
- **Geometry** (32 features), associated to the strokes or signature aspect-ratio: 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.

Feature selection on this 100-feature set is performed using the SFFS algorithm [10], which is set to minimize the system EER using a classifier based on the Mahalanobis distance.

4. Experimental Setup

A subset of the PDA and pen tablet signature corpus of the BioSecure multimodal biometric database [6] is used for experiments. It consists of 120 users, with 20 genuine signatures and 20 skilled forgeries per user and acquisition device (PDA and pen tablet). The genuine signatures were acquired in two different sessions separated by an average period of two months, being 5 signatures from the first session and the remaining 15 from the second session. In each session, signatures were produced by the user in blocks of 5, leaving a gap of some minutes between each block. Signatures were captured with a PDA while the user was standing and holding the PDA with one hand in the handheld scenario, whereas for the pen tablet case, they were captured while the user was sitting, using a pen on a paper

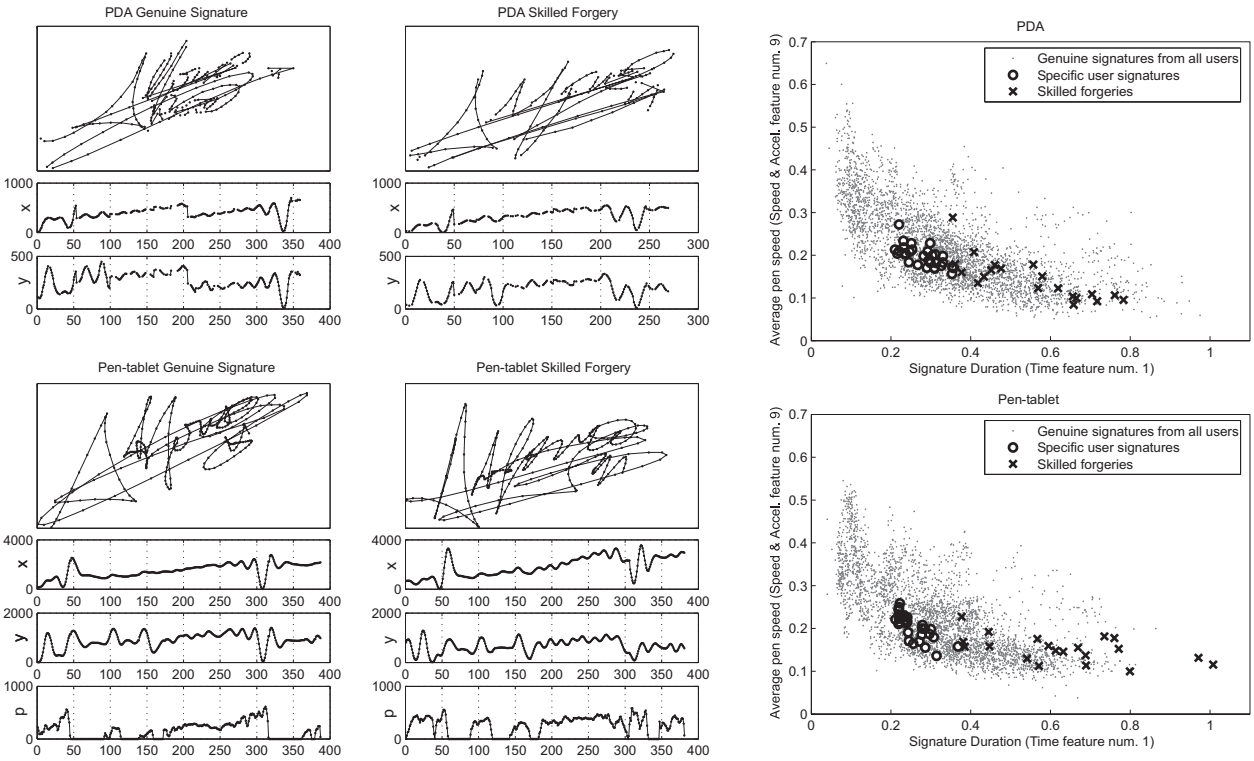

Figure 2. Left: signatures from a user of the database on both scenarios and their corresponding signals used for the experiments. No pressure signals are available for the signatures captured with PDA. Right: distribution of two example features for a particular user.

placed over the tablet. This emulates real operating conditions.

Only the x and y position signals and the sample timestamps are captured by the PDA, while pressure information and pen orientation is also provided by the pen tablet. Skilled forgeries for each user were performed by 4 different users (5 forgeries each) in a "worst case" scenario: each forger had visual access to the dynamics of the genuine signature and a tracker tool allowing to see the original strokes in both scenarios. An example of the captured signatures, their associated signals and the distribution of two features on both scenarios is shown in Fig. 2. It can be seen that signatures captured with the PDA have missing samples due to capture errors.

From each subset, the signatures from 50 users are used for development purposes, while the remaining 70 will be used to compute the verification performance. The 5 signatures from the first session are used to compute the user models. This setup follows the protocol of the BIOSECURE Multimodal Evaluation Campaign (BMEC), where a subset of 50 users was previously released for algorithm tuning before submission to the competition, which was performed by the evaluation organizers on sequestered test data. For the PDA subset, a preprocessing step is performed to interpolate missing samples.

Random forgery scores (the case where a forger uses his own signature claiming to be another user of the system) are obtained by comparing the user model to one signature sample of all the remaining users. Skilled forgery scores are computed by comparing all of the 20 available skilled forgeries per user with its own model.

The experiments are structured as follows: first, an analysis based on Fisher Discriminant Ratio (FDR) for each individual feature is performed over the development set. Next, feature selection based on the SFFS algorithm is performed (separately for random and skilled forgeries) to obtain an optimum feature subset for the handheld and tablet scenario. The contribution of each type of features (Time, Speed, etc.) in the optimum feature set for each scenario is then studied. Finally, the verification performance of the optimum feature subset over the test set (the remaining 70 users) is studied, and fusion with an HMM-based system is performed.

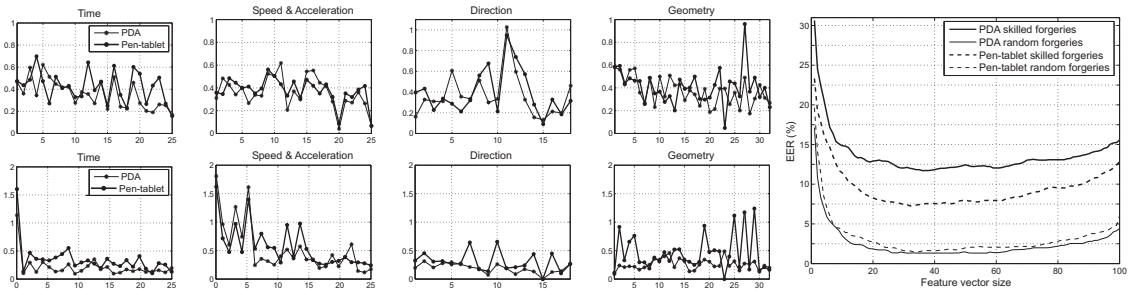

Figure 3. Median FDR over the development set for random (top) and skilled forgeries (bottom). Right: EER vs. number of features selected by the SFFS algorithm on the development set.

5. Results

To analyze the discriminative power embedded in the different feature types, the Fisher Discriminant Ratio is used as in [8], but computed for individual features and individual subjects. The median FDR (as the mean value is affected by outliers) over the different users in the development user set is depicted in Fig. 3 for the four feature types specified in Sect. 3. The FDR provides an intuitive measure of discriminative power, as it increases with the inter-class variability and decreases with the intra-class variability. The median FDR is computed differently for random and skilled forgeries. In the case of random forgeries, for each user, the FDR between the user samples and the rest of the genuine signatures in the database is computed, while for skilled forgeries, the FDR is computed between the genuine signatures of the user and the available skilled forgeries.

From Fig. 3, we observe that the median FDR for each feature is similar in the pen tablet and the PDA scenario when random forgeries are considered (top row). On the contrary, it is higher for pen tablet than PDA in the case of skilled forgeries (bottom row). This suggests that the verification performance in the PDA scenario against skilled forgeries would be *a priori* lower than for pen tablet independently from the classifier used. In Fig. 2, the distribution of the normalized values of two example features is depicted.

In Fig. 3 (right) the evolution of the system EER according to the size of the optimum feature vector selected by the SFFS algorithm is depicted. It can be observed that while the behavior for the case of random forgeries is similar on both scenarios, the verification performance is significantly better for skilled forgeries in the pen-tablet scenario.

The contribution of each type of feature is analyzed in Fig. 4. An histogram of each type of feature for different sizes of the optimal feature vector computed by the SFFS algorithm is depicted for random and skilled forgeries on both scenarios. As can be seen in Fig. 4(a) and (b), Geometry features represent a very high proportion in the PDA scenario, with a much reduced contribution of the rest of features. On the contrary, in the pen tablet scenario (Fig. 4(c) and (d)), the contribution of Geometry features is balanced with the one of Time and Speed and Acceleration features.

These results reveal that for the PDA scenario, the discriminative power of dynamic features such as Speed and Acceleration and Time features may be much lower than geometrical features. Thus, for the case of skilled forgeries, the verification performance is degraded on the PDA scenario, as Geometry features are commonly the easiest to forge.

Fusion of the PDA global feature system optimized for skilled forgeries with an user-dependent HMM system is performed via weighed sum of the match scores. The fusion weights have been heuristically adjusted to optimize both the random and skilled EERs. An optimum vector size of 50 features is selected. In the HMM system, the number of states is proportional to the mean length of the user training signatures, and the number of Gaussian Mixtures in the observations is set to maximize the likelihood of the training data with a limit of 32 mixtures. This HMM system is based on the one presented in [5]. The verification results are shown in the first row of Table 1 for the test set (70 users).

5.1. BioSecure Evaluation

The BioSecure Multimodal Evaluation Campaign [6] was held in 2007 and was composed of an Access Control Scenario and a Mobile Scenario. A signature verification modality was present in the Mobile Scenario, where signatures from the BioSecure multimodal database were used for the evaluation. A total of 11 systems were presented, from 6 independent research groups. The evaluation protocol is equivalent to the one followed in this paper (with another

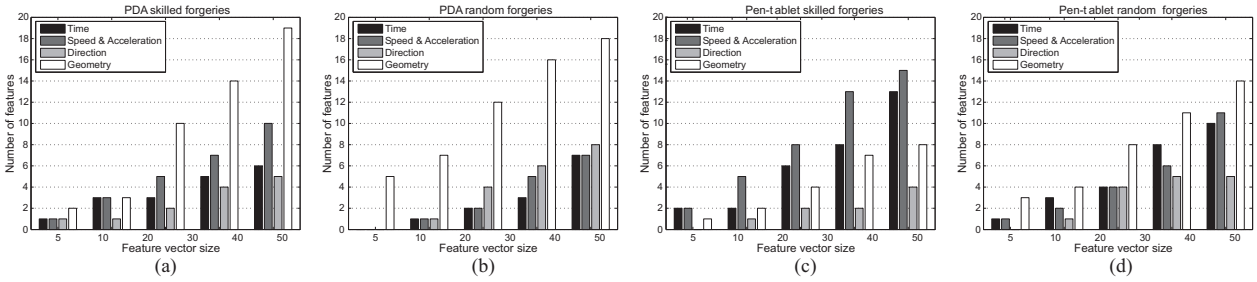

Figure 4. Histograms of feature types for different optimum feature-vector sizes and scenarios.

Table 1. PDA scenario EER comparison for random (rd) and skilled (sk) forgeries.

System	EER_{rd}	EER_{sk}
Proposed system (HMM+features)	4.0%	11.9%
BMEC best for sk. forgeries [11]	8.07%	13.43%
BMEC best for rd. forgeries [8]	4.03%	13.58%

data subset, captured in the same conditions). The best results of the evaluated systems for random and skilled forgeries are shown in Table 1, as well as the performance of the system proposed in this work. The winner system against skilled forgeries was based on an ensemble of local and global Gaussian Mixture Models and derived from [8]. The best system for random forgeries was HMM-based, using fusion of the likelihood and Viterbi path scores [11]. As can be seen, a notable verification performance has been obtained in the present work compared to those systems.

6. Conclusions and Future Work

The importance of adapting the traditional tablet-based signature verification systems to the new PDA scenario has been stated. The observed low discriminative power of dynamic features (time, speed and acceleration) in the PDA scenario suggests that ergonomics and an unfamiliar surface and signing device (touch-screen and PDA stylus vs. traditional pen and paper) may be affecting the signature process. On the other hand, the users are still able to reproduce the geometry of their own signature, which is shown by the higher consistency of geometric features. Future work includes the application of techniques aimed to compensate the increased variability found in the handheld scenario, like feature subset transformations based on session-invariant subspaces, recently introduced with significant success in the speaker recognition literature [12].

References

[1] R. Plamondon and G. Lorette. Automatic signature verification and writer identification: the state of the art. *Pattern Recognition*, 22(2):107–131, 1989.

[2] J. Fierrez and J. Ortega-Garcia. *Handbook of Biometrics*, chapter On-line signature verification. Eds. A. K. Jain, A. Ross and P. Flynn, Springer, 2008.

[3] L. L. Lee et al. Reliable on-line human signature verification systems. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(6):643–647, 1996.

[4] A. Kholmatov and B. Yanikoglu. Identity authentication using improved online signature verification method. *Pattern Recognition Letters*, 26(15):2400–2408, 2005.

[5] J. Fierrez, D. Ramos-Castro, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. HMM-based on-line signature verification: feature extraction and signature modeling. *Pattern Recognition Letters*, 28(16):2325–2334, 2007.

[6] F. Alonso-Fernandez et al. Dealing with sensor interoperability in multi-biometrics: the UPM experience at the BioSecure Multimodal Evaluation 2007. In *Defense and Security Symposium, Proc. SPIE*, USA, 2008.

[7] D. Y. Yeung et al. SVC2004: First International Signature Verification Competition. In *Proc. ICBA*, pages 16–22. Springer LNCS-3072, 2004.

[8] J. Richiardi et al. Local and global feature selection for on-line signature verification. In *Proc. ICDAR*, Seoul, Korea, August-September 2005.

[9] J. Fierrez-Aguilar et al. An on-line signature verification system based on fusion of local and global information. In *Proc. AVBPA*, pages 523–532. Springer LNCS, 2005.

[10] A. K. Jain and D. Zongker. Feature selection: evaluation, application, and small sample performance. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(2):153–158, 1997.

[11] B. L. Van et al. On using the viterbi path along with HMM likelihood information for online signature verification. *IEEE Trans. on Systems, Man, and Cybernetics, Part B*, 37(5):1237 – 1247, 2007.

[12] P. Kenny et al. Speaker and session variability in GMM-based speaker verification. *IEEE Trans. on Audio, Speech and Language Processing*, 15(4):1448–1460, 2007.