Mobile signature verification: feature robustness and performance comparison

Marcos Martinez-Diaz, Julian Fierrez, Ram P. Krish, Javier Galbally

Biometric Recognition Group - ATVS, Escuela Politecnica Superior, Universidad Autonoma de Madrid,
Campus de Cantoblanco, C/ Francisco Tomas y Valiente, 11, 28049 Madrid, Spain
E-mail: marcos.martinez@uam.es

Abstract: In this study, the effects of using handheld devices on the performance of automatic signature verification systems are studied. The authors compare the discriminative power of global and local signature features between mobile devices and pen tablets, which are the prevalent acquisition device in the research literature. Individual feature discriminant ratios and feature selection techniques are used for comparison. Experiments are conducted on standard signature benchmark databases (BioSecure database) and a state-of-the-art device (Samsung Galaxy Note). Results show a decrease in the feature discriminative power and a higher verification error rate on handheld devices. It is found that one of the main causes of performance degradation on handheld devices is the absence of pen-up trajectory information (i.e. data acquired when the pen tip is not in contact with the writing surface).

1 Introduction

Signature verification is still a challenging task within biometrics. Owing to their behavioural nature (as opposed to anatomic biometric traits), signatures present a notable variability even between successive realisations, which can be increased over medium or large periods of time [1]. Moreover, evaluating the robustness of a system against forgeries is complex, as highly skilled forgers are rarely available during the collection of research databases. 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. Reliable automatic signature verification is nevertheless an active research field [2] because of the widespread social and legal acceptance of signatures as a validation means.

On-line or dynamic signature verification systems use discrete-time functions sampled from the pen tip motion (e.g. x and y coordinates) to perform authentication. These signals may be captured, for example, with pen tablets or touch screens. Dynamic systems have reached traditionally a better verification performance than off-line systems, which consider only the static handwritten signature image, since more levels of information than the signature still image are available [3].

Dynamic signature verification systems can be classified into two main categories. 'Feature-based' or global systems, which model the signature as a holistic multidimensional vector composed of global features such as average pen speed or number of pen-ups [4], and 'function-based' or local systems that perform signature matching using the captured discrete-time functions (pen coordinates, pressure

and so on) [5]. Feature-based systems use statistical classifiers such as Parzen-windows or Gaussian mixture models [6, 7], whereas function-based systems traditionally use dynamic time warping (DTW) [8, 9] or hidden Markov models (HMM) among other techniques [10].

Smartphones and handheld devices have recently gathered a high level of popularity in the context of ubiquitous access to information and services. These devices represent an attractive target for the deployment of a signature verification system, providing enough processing capabilities and a touch-based interface [11]. However, signature verification on handheld devices is affected by factors not present in other input devices primarily because of a small input area, poor ergonomics or the fact that the user may be in movement. Users must sign on an unfamiliar and usually unstable surface with a small stylus or a finger. As a consequence, the signature generation process may be degraded.

The BioSecure Signature Evaluation Campaign (BSEC 2009) [12], with the participation of several independent research institutions, has shown that the performance of signature verification using samples captured on a handheld device is significantly lower than with signatures captured on a pen tablet. Nevertheless, the impact of handheld devices on local and global signature features has not been systematically studied to the extent of our knowledge. A preliminary statistical comparison of such signature features acquired with several devices, including a pen tablet and a personal digital assistant (PDA) was performed in [13], showing that there are significant differences in feature distributions among different devices.

The objective of this work is to study the effects of mobile acquisition conditions in automatic signature verification. We focus on the impact of mobile conditions on the feature

discriminative power of different types of features (local and global) compared to the traditional pen tablet scenario using discriminant analysis of individual features and feature selection algorithms. The performance of state-of-the-art verification systems is also studied in both scenarios, using a global and an HMM-based local system. Two feature sets are considered in this work. A global feature set based on the one described in [14], and a local set which contains most local features proposed in recent years for dynamic signature verification.

Two different databases are used in the experiments: (i) the BioSecure Multimodal Database (BMDB), as a standard benchmark [15]; (ii) a signature database captured using a state-of-the-art device (Samsung Galaxy Note). The BMDB signature database has two subcorpora, one captured on a PDA and other on a digitising pen tablet. They correspond to the same users in both devices, allowing a fair comparison between them.

This paper is structured as follows. First, related work on signature verification on handheld devices is reviewed in Section 2. In Section 3, the global and local features considered in this work are described and a global and a local verification system are also presented. The experimental protocol and databases are described in Section 4. Results are reported in Section 5. Conclusions are finally drawn in Section 6.

2 Related work

Little research has been carried out in the field of dynamic signature verification on handheld devices. In most works related to automatic signature verification, signature databases are captured using a pen tablet [2]. As a matter of fact, most research-oriented signature databases have been acquired with a pen tablet [16], although there is an emerging interest in signature-based authentication using alternative devices [17].

Compared to touch screens on PDAs or handheld devices, pen tablets usually capture more information than the pen trajectory, namely pen orientation (azimuth and altitude) and pen pressure. Moreover, pen tablets also detect the pen trajectory when the tip is not in contact with the surface, allowing trajectory acquisition pen-ups.

The BMDB [15] contains, among other biometric traits, two signature datasets from the same set of donors. One dataset was captured with a pen tablet (DS2 dataset) and another with a PDA (DS3 dataset). In Fig. 1, the capture conditions of both datasets are shown. In 2007, the BioSecure Multimodal

Evaluation was held, where verification algorithms from several European research institutions were compared using the PDA dataset [18]. It was found that error rates were notably higher than in previous competitions, such as SVC 2004 [19], where signatures had been captured on a pen tablet. In 2009, the BSEC was aimed towards comparing the verification performance between the handheld scenario and the pen tablet scenario [12]. Two different tasks were reported. In Task 1, a direct comparison of verification performance using a pen tablet against a PDA for signature acquisition was carried out, with signatures from the BMDB database. Task 2 studied the verification performance variation with respect to the information content in signatures [20]. Results of Task 1 showed that the participating signature verification algorithms had a significant lower performance against skilled forgeries when signatures were captured on a PDA compared to a pen tablet. On the other hand, verification performance against random forgeries was less negatively affected in the PDA scenario.

It is also known from previous works that features extracted from signatures acquired with different devices present statistical distributions that might be significantly different [13]. These statistical differences between features from different devices may affect device interoperability and may also result in large verification performance differences among sensors. In [21], the authors compare the error rates of two systems when signatures are captured with two different tablet-PCs. It is shown that the performance depends on the sampling quality of the device used for enrolment. In [22], the effects of constraining the available signing space are studied, although not specifically for handheld devices. The authors show that the lack of space affects signature complexity, may cause hesitation marks, and reduce fluency, among other factors.

Signatures captured with a pen tablet and a handheld device have also been compared from the point of view of their entropy or information content. In [20] a client-entropy measure is defined, and it is shown that signatures captured with a PDA have a higher entropy than those captured with a pen tablet. The entropy measure defined in that work increases in general with signature variability and graphically simple signatures. Higher verification error rates for signatures with higher entropy are reported.

3 Features and recognition systems

The objective of this work is to study the effects of handheld devices on signature features and verification performance.
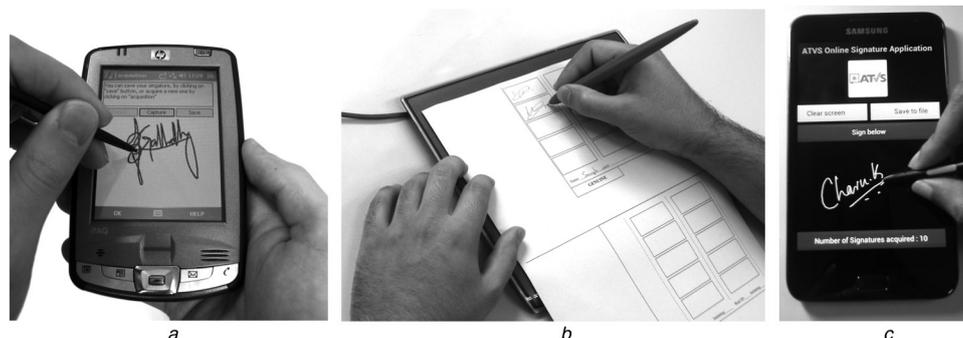
Fig. 1 *Signature capture conditions on different devices*
a PDA signature capture process in the BIOSECURE DS3 – mobile scenario dataset
b Pen tablet capture process in the BIOSECURE DS2 – access control scenario dataset
c Signature capture process on a mobile device (Samsung Galaxy Note) used for validation experiments

Two totally different recognition systems have been selected, one based on global features and the other based on local features. In order to reach these goals, two signatures datasets have been used, containing samples acquired on a pen tablet and a handheld device. In Fig. 2, the experimental approach that has been followed is depicted.

3.1 Global features

Feature-based or global signature verification systems have been extensively studied in the past [4, 14, 23]. Signatures are described in this case by an *n*-dimensional vector, containing features related to shape, and timing-events among other feature types. In this work, a large set of 100 global features is considered, which comprises a high proportion of the best performing global features studied in the literature [4, 24, 25]. A complete description of the feature set is shown in Table 1.

Features are extracted directly from the pen motion discrete-time signals and are normalised between [0, 1] using *tanh*-estimators [26]. These global features can be divided into four categories according to the physical magnitude measured:

- *Time (25 features):* related to signature duration, or timing of events such as pen-ups or local maxima.
- *Speed and acceleration (25 features):* obtained from the first- and second-order time derivatives of the position time functions.
- *Direction (18 features):* extracted from the path trajectory. Examples are starting direction or mean direction between pen-ups.
- *Geometry (32 features):* associated with the strokes or signature aspect-ratio.

3.1.1 Global verification system: A classifier based on a simplified version of the Mahalanobis distance is used in this work, in order to compare an input signature with a claimed user model. This distance measure has the advantage of being relatively simple to compute and generic enough to provide a reasonable empirical estimate of the statistical class separability achieved by the feature vectors. User models $C = (\mu, \Sigma)$ are created from a training set of genuine signatures, where μ and Σ are the mean vector and covariance matrix obtained from the training signatures. A diagonal covariance matrix is used, and the values below a fixed threshold are replaced by the threshold value. This is done to avoid obtaining a singular covariance matrix because of the limited number of training samples in comparison to the problem dimensionality, and to simplify

the implementation of this algorithm in handheld devices with limited processing power. The threshold value is 0.00085 and has been heuristically obtained in preliminary experiments. Thus, the matching score *s* is obtained as the inverse of the 'simplified' Mahalanobis distance between the input signature feature vector *x* and the claimed user model *C*

$$s(x, C) = \left((x - \mu)^{\mathrm{T}} (\Sigma)^{-1} (x - \mu) \right)^{-1/2} \tag{1}$$

If the score $s(x, C)$ is above a specific threshold, the signature is considered as genuine. In contrast it is rejected by the system.

3.2 Local features

The local feature set considered in this work is an extension of the feature set described in [10], which has been extended by adapting features from [23, 27, 28]. In the original set, seven discrete-time functions are extracted from the pen tip trajectory and the pen pressure, from which the first- and second-order derivatives are computed, leading to a 21-dimensional feature vector [10]. In this work, all second-order derivatives except those extracted from *x* and *y* coordinates are discarded since they showed a very low contribution in the verification performance (as corroborated in [23]). Thus, 16 functions from the original set are used (7 + 7 derivatives + 2 second-order derivatives), which correspond to features numbered 1–14 and 17–18 in Table 2. The set has been extended with 11 functions extracted from [23, 27, 28].

In Table 2 the resulting set of 27 functions is described. As in the case of global features, this feature set comprises a high proportion of the features proposed in the recent literature for local signature verification.

This feature set assumes the availability of pressure and pen-inclination information, although that is not usually the case for signatures captured with a handheld device. In those cases, only 21 features can be extracted from the raw signals (see caption of Table 2).

3.2.1 Local verification system: An HMM system is used in the experiments. This system is based on the one described in [10], which reached second position in Task 2 of the Signature Verification Competition 2004 [19]. Each user is modelled with a 2-state HMM with 32 Gaussian mixtures per state. Similarity scores are computed as the log-likelihood of the target signature (using the Viterbi algorithm) divided by the total number of samples of the
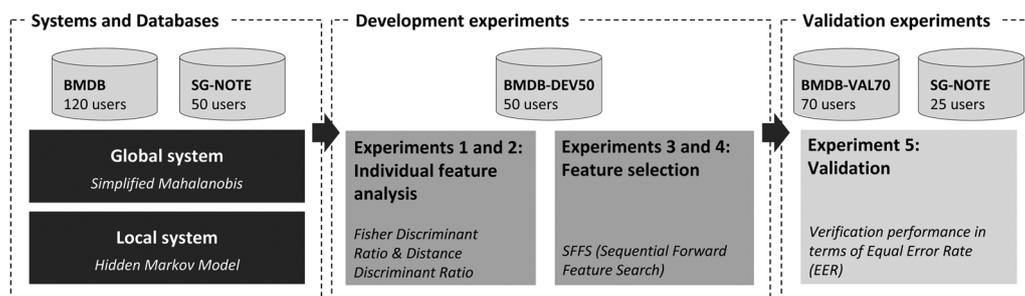
Fig. 2 *Diagram of the experimental setup followed in this work*
Global and local systems are presented in Section 3. The experimental protocol and databases are described in Section 4. Results are reported in Section 5

Table 1 Global feature set considered in this contribution

#	Time related feature						#	Direction related feature					
#	Speed and Acceleration related feature						#	Geometry related feature					

#	Feature description	Ps	Pr	Tr	Ts	Ur	Us	#	Feature description	Ps	Pr	Tr	Ts	Ur	Us
1	signature total duration T_s	✓	✓	✓	✓	✓	✓	2	(pen-down duration T_w)/T_s						
3	(1st $t(v_{max})$)/T_w							4	$T(v_x > 0)/T_w$		✓	✓	✓	✓	✓
5	$T(v_x < 0)/T_w$			✓				6	$T(v_y > 0)/T_w$	✓					
7	$T(v_y < 0)/T_w$				✓			8	$T(v_x > 0|pen-up)/T_w$	✓					
9	$T(v < 0|pen-up)/T_w$				✓	✓		10	$T(v_y > 0|pen-up)/T_w$	✓					✓
11	$T(v_x < y|pen-up)/T_w$							12	$T(1stpen-up)/T_w$				✓		✓
13	$T(2ndpen-up)/T_w$	✓			✓			14	$T(2nd\ pen-down)/T_s$				✓		
15	$T(3rdpen-down)/T_s$	✓		✓	✓			16	(1st $t(v_{y,max})$)/T_w	✓	✓	✓		✓	
17	(1st $t(v_{y,min})$)/T_w		✓	✓	✓	✓		18	(1st $t(v_{x,max})$)/T_w			✓	✓	✓	
19	(1st $t(v_{x,min})$)/T_w	✓	✓	✓	✓	✓		20	$T((dy/dt)/(dx/dt) > 0)/T((dy/dt)/(dx/dt) < 0)$			✓			
21	$T(curvature > threshold_{curv})/T_w$							22	(1st $t(x_{max})$)/T_w				✓		✓
23	(2nd $t(x_{max})$)/T_w	✓		✓			✓	24	(3rd $t(x_{max})$)/T_w		✓				
25	(2nd $t(y_{max})$)/T_w	✓			✓		✓	26	(3rd $t(y_{max})$)/T_w			✓		✓	
27	(average velocity $\bar{v}$)/v_{max}			✓	✓			28	$N(v_x = 0)$			✓	✓		✓
29	$N(v_y = 0)$	✓	✓	✓		✓	✓	30	$\bar{v}/v_{y,max}$	✓	✓	✓	✓		✓
31	$\bar{v}/v_{y,max}$				✓			32	(velocity rms v)/v_{max}	✓		✓	✓		✓
33	(centripetal acceleration rms a_c)/a_{max}	✓				✓		34	(tangential acceleration rms a_t)/a_{max}	✓		✓	✓		✓
35	(acceleration rms a)/a_{max}				✓		✓	36	(integrated abs. centr. acc. a_{Ic})/a_{max}	✓	✓		✓	✓	✓
37	(velocity correlation $v_{x,y}$)/v^2_{max}	✓	✓			✓	✓	38	standard deviation of v_x					✓	
39	standard deviation of v_y			✓	✓		✓	40	standard deviation of a_x						
41	standard deviation of a_y			✓	✓	✓	✓	42	average jerk			✓	✓		
43	$\bar{J}_x$							44	$\bar{J}_y$						
45	j_{max}	✓	✓		✓			46	$j_{x,max}$			✓			✓
47	$j_{y,max}$				✓	✓	✓	48	j_{rms}			✓			✓
49	$t(j_{max})/T_w$		✓				✓	50	$t(j_{x,max})/T_w$		✓	✓	✓	✓	✓
51	$t(j_{y,max})/T_w$			✓	✓	✓	✓	52	$N(pen-ups)$	✓	✓	✓		✓	✓
53	$N(sign\ changes\ of\ dx/dt\ and\ dy/dt)$	✓	✓	✓	✓	✓	✓	54	$T((dx/dt)(dy/dt) > 0)/T((dx/dt)(dy/dt) < 0)$						
55	$\theta(initial\ direction)$		✓					56	$\theta(1st\ to\ 2nd\ pen-down)$			✓	✓	✓	✓
57	$\theta(1st\ pen-down\ to\ 1st\ pen-up)$	✓	✓					58	$\theta(1st\ pen-down\ to\ 2nd\ pen-up)$					✓	
59	$\theta(2nd\ pen-down\ to\ 2nd\ pen-up)$		✓	✓	✓		✓	60	$\theta(before\ lastpen-up)$	✓	✓				
61	$\theta(1st\ pen-down\ to\ last\ pen-up)$		✓					62	direction histogram s_1	✓	✓			✓	✓
63	direction histogram s_2	✓	✓	✓	✓		✓	64	direction histogram s_3						
65	direction histogram s_4	✓		✓	✓	✓	✓	66	direction histogram s_5			✓	✓		
67	direction histogram s_6							68	direction histogram s_7						
69	direction histogram s_8	✓	✓					70	direction change histogram c_2				✓	✓	
71	direction change histogram c_3		✓					72	direction change histogram c_4	✓					
73	$A_{min} = (y_{max} - y_{min})(x_{max} - x_{min})\ \left(\Delta_x = \sum_{i=1}^{pen-downs} (x_{max|i} - x_{min|i})\right)\Delta_y$	✓				✓	✓	74	(max distance between points)/A_{min}						
75	$(x_{1st\ pen-down} - x_{max})/\Delta_x$		✓	✓	✓	✓		76	$(x_{1st\ pen-down} - x_{min})/\Delta_x$			✓	✓	✓	✓
77	$(x_{last\ pen-up} - x_{max})/\Delta_x$	✓	✓					78	$(x_{last\ pen-up} - x_{min})/\Delta_x$	✓					
79	$(y_{1st\ pen-down} - y_{max})/\Delta_y$		✓			✓		80	$(y_{1st\ pen-down} - y_{min})/\Delta_y$	✓		✓		✓	✓
81	$(y_{last\ pen-up} - y_{max})/\Delta_y$	✓			✓	✓	✓	82	$(y_{last\ pen-up} - y_{min})/\Delta_y$	✓					
83	$(x_{max} - x_{min})\Delta_y/(y_{max} - y_{min})\Delta_x$	✓	✓					84	(standard deviation of x)/Δ_x	✓		✓		✓	
85	(standard deviation of y)/Δ_y	✓						86	$(T_w \bar{v})/(y_{max} - y_{min})$		✓	✓		✓	
87	$(T_w \bar{v})/(y_{max} - y_{min})$		✓	✓				88	$(x_{max} - x_{min})/x_{acquisition\ range}$	✓			✓	✓	✓
89	$(y_{max} - y_{min})/y_{acquisition\ range}$							90	$(\bar{x} - x_{min})/\bar{x}$		✓				
91	spatial histogram t_1		✓	✓		✓		92	spatial histogram t_2			✓	✓	✓	✓
93	spatial histogram t_3	✓	✓					94	spatial histogram t_4		✓				
95	$N(local\ maxima\ in\ x)$	✓	✓			✓	✓	96	$(x_{2nd\ local\ max} - x_{1st\ pen-down})/\Delta_x$		✓				
97	$(x_{3rd\ local\ max} - x_{1st\ pen-down})/\Delta_x$			✓				98	$N(local\ maxima\ in\ y)$	✓	✓	✓	✓		✓
99	$(y_{2nd\ local\ max} - y_{1st\ pen-down})/\Delta_y$							100	$(y_{3rd\ local\ max} - y_{1st\ pen-down})/\Delta_y$	✓	✓	✓	✓		

T denotes time interval, t denotes time instant, N denotes number of events, and θ denotes angle. All notations are defined or referenced in the table. Features 36, 37, 62 and 91 are based on [25]. The optimal 40-feature subsets, as described in the Experimental Results (Section 5.2), are shown for each optimisation scenario: Ps and Pr denote PDA skilled and random forgeries, Ts and Tr pen tablet skilled and random forgeries and Us and Ur refer to pen tablet with interpolated pen-ups against skilled and random forgeries, respectively

signature signal. In order to keep scores between a reasonable range, normalised scores $\hat{s}$ between (0,1) are obtained as $\hat{s} = \exp(s(\mathbf{x}, \mathbf{C})/30)$, where $s(\mathbf{x}, \mathbf{C})$ is the score returned by the HMM system and $\mathbf{x}$ and $\mathbf{C}$ represent, respectively, the input signature and the enrolled model of the claimed identity.

Table 2 Local feature set presented in this contribution

#	Feature	Description
1	x-coordinate	x_n
2	y-coordinate	y_n
3	pen-pressure	z_n
4	path-tangent angle	$\theta_n = \arctan(\dot{y}_n/\dot{x}_n)$
5	path velocity magnitude	$v_n = \sqrt{\dot{y}_n^2 + \dot{x}_n^2}$
6	log curvature radius	$\rho_n = \log(1/\kappa_n) = \log(v_n/\dot{\theta}_n)$, where κ_n is the curvature of the position trajectory
7	total acceleration magnitude	$a_n = \sqrt{t_n^2 + c_n^2} = \sqrt{\dot{v}_n^2 + v_n^2 \dot{\theta}_n^2}$, where t_n and c_n are, respectively, the tangential and centripetal acceleration components of the pen motion
8–14	first-order derivative of features 1–7	$\dot{x}_n, \dot{y}_n, \dot{z}_n, \dot{\theta}_n, \dot{v}_n, \dot{\rho}_n, \dot{a}_n$
15	pen azimuth	γ_n
16	pen altitude	ϕ_n
17–18	first-order derivative of features 15–16	$\dot{\gamma}_n, \dot{\varphi}_n$
19–20	second-order derivative of features 1–2	$\ddot{x}_n, \ddot{y}_n$
21	ratio of the minimum over the maximum speed over a window of 5 samples	$v_n^r = \min\{v_{n-4}, \ldots, v_n\}/\max\{v_{n-4}, \ldots, v_n\}$
22–23	angle of consecutive samples and first-order difference	$\alpha_n = \arctan(y_n - y_{n-1}/x_n - x_{n-1})$ $\dot{\alpha}_n$
24	sine	$s_n = \sin(\alpha_n)$
25	cosine	$c_n = \cos(\alpha_n)$
26	stroke length to width ratio over a window of 5 samples	$r_n^5 = \dfrac{\sum_{k=n-4}^{k=n} \sqrt{(x_k - x_{k-1})^2 + (y_k - y_{k-1})^2}}{\max\{x_{n-4}, \ldots, x_n\} - \min\{x_{n-4}, \ldots, x_n\}}$
27	stroke length to width ratio over a window of 7 samples	$r_n^7 = \dfrac{\sum_{k=n-6}^{k=n} \sqrt{(x_k - x_{k-1})^2 + (y_k - y_{k-1})^2}}{\max\{x_{n-6}, \ldots, x_n\} - \min\{x_{n-6}, \ldots, x_n\}}$

Upper dot notation (e.g. $\dot{x}_n$) indicates time derivative, and the subindexes (integers) indicate time sampling instants. Features 3, 10, 15, 16, 17 and 18 are not available in the mobile scenario, as the pressure and pen-inclination information is not acquired by this type of devices

4 Experimental protocol

4.1 Databases

Two databases are used in the experiments, the BMDB, acquired using a pen-tablet and a PDA [15], and a database captured using a Samsung Galaxy Note device, referred to as SG-NOTE.

A subset of 120 users from the BMDB is used in this work [available at the BioSecure Foundation web site: http://biosecure.it-sudparis.eu/AB/index.php?option=com_content&view=article&id=72 as the '120 common DS2/DS3' signature dataset]. It contains 20 genuine signatures and 20 skilled forgeries per user and acquisition device (PDA and pen tablet). Genuine signatures were acquired in two different sessions separated by an average period of two months. The first five signatures correspond to the initial session whereas the remaining 15 belong to the second session. Signatures were captured with a PDA while the user was standing and holding the device with one hand in the handheld scenario, whereas for the pen tablet case they were acquired while the user was sitting, using a pen on a paper placed over the tablet (see Fig. 1a and b). This emulates real operating conditions.

In both devices, skilled forgeries for each user were performed by four different forgers (five forgeries each) under 'worst case' conditions: each forger had visual access to the dynamics of the genuine signature using a tracker tool that allowed replaying the original strokes.

Only the x and y position signals and the sample timestamps are captured by the PDA, whereas pressure (z) and pen orientation (θ, γ) signals are also acquired by the pen tablet. Pen trajectories during pen-ups (when the pen tip is not in contact with the tablet surface) are recorded by the pen tablet but are not available in the PDA dataset. It is found in the pen tablet dataset that, for each genuine signature, an average of 18% of sampled points correspond to pen-up trajectories (i.e. when the pen tip is not in contact with the tablet surface). A histogram of the proportion of sample points during pen-ups compared to the total signature samples is depicted in Fig. 3. In order to evaluate the effect of the lack of pressure and inclination information and pen motion during pen-ups, a third signature dataset is artificially created by removing the samples produced during pen-ups (i.e. having pressure values equal to 0) in the pen tablet dataset. This set will be referred to as 'tablet interpolated pen-ups'. Pen-up trajectories are interpolated in the PDA and in the tablet interpolated pen-ups dataset. Cubic splines are selected for interpolation as they provided a better verification performance in preliminary experiments, which are omitted for the sake of clarity. For the PDA subset, an additional preprocessing step is performed to interpolate erroneous (missing) samples.

From each of the three BMDB subsets (i.e. PDA, tablet and tablet interpolated pen-ups), each one containing 120 users, signatures from the first 50 users are used for development purposes (i.e. individual feature analysis and feature selection), whereas the remaining 70 are left to validate the performance of the optimal feature vectors selected by the Sequential Forward Floating Search (SFFS) algorithm. We will refer to the development datasets as BMDB-DEV50 and to the validation datasets as BMDB-VAL70.

This setup follows the protocol of the BSEC [12], where a subset of 50 users was released for algorithm tuning prior to the competition, which was later carried out using a different test dataset.
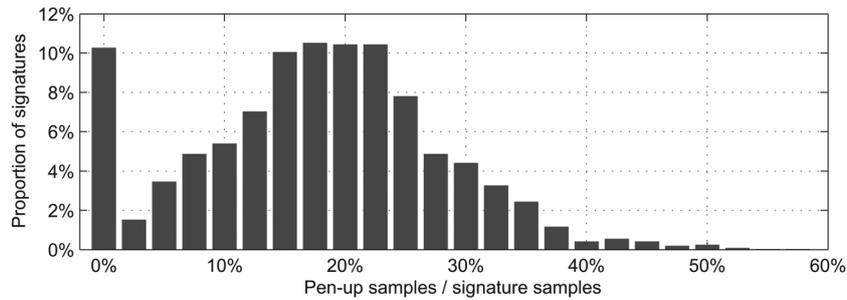
Fig. 3 *Histogram of signatures classified by the proportion of sampled points during pen-up trajectories against total signature sample points, computed on the pen tablet signature dataset*

The SG-NOTE database [available at http://atvs.ii.uam.es] is also used for performance validation, in addition to the BMDB-VAL70 subset. This dataset was captured by the authors using a Samsung Galaxy Note mobile phone and contains signatures from 25 users. The SG-NOTE database was captured in two different sessions with an average gap of 5 days between them. In each session, signatures were acquired in two blocks of five samples, with a short break between blocks. No skilled forgeries are available in this database. Consequently, the database contains a total amount of 500 signatures (25 users × 2 sessions × 10 signatures per session). See Fig. 1*c* for an example signature acquisition in SG-NOTE.

The five genuine signatures from the initial session are used for enrolment, both for the global and local systems. 'genuine' user scores are computed using the remaining from the second session (15 signatures in BMDB and 10 signatures in SG-NOTE). 'Random forgery' scores (the case where a forger uses his own signature claiming to be a different user) are obtained by comparing the user model to one signature sample of all the remaining users. 'Skilled forgery' scores for the BMDB datasets are computed comparing the 20 available skilled forgeries per user with his or her own model (trained with five signatures, as stated before).

4.2 Development and validation experiments

The experiments are structured as follows: first, a global and local individual feature analysis is performed on signatures from the BMDB-DEV50 development dataset (Experiments 1 and 2). Optimal feature combinations are then computed using feature selection (Experiments 3 and 4). Finally, results are validated using the BMDB-VAL70 and SG-NOTE datasets (Experiment 5).

4.2.1 Experiment 1 – global feature analysis: The discriminative power of global features can be measured using the Fisher discriminant ratio (FDR) for each individual feature. 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 FDR D for the ith feature from user u is computed as follows

$$D_i(u) = \frac{(\mu_{G_i} - \mu_{F_i})^2}{\sigma_{G_i}^2 + \sigma_{F_i}^2} \qquad (2)$$

where μ and σ are the average and standard deviation, respectively, of the genuine signature sample set G_i and the forged sample set F_i. We use this measure in this work to

compare the discriminative power of each feature defined in Table 1 between the mobile and the pen tablet scenario.

4.2.2 Experiment 2 – local feature analysis: In contrast to the case of global features, the application of the FDR to compute the discriminative power of individual local features is impractical. This is because of the fact that local features are time functions. As a consequence, the computation of distances between average feature values as defined in the FDR does not represent a realistic measure.

A distance-based discriminative measure using time functions is proposed in [27] to overcome this limitation. In that work, a consistency value is described, which provides a similar measure to the FDR at least from an intuitive point of view, as it decreases when genuine features are far apart among them and close to forgeries and vice versa. We use the DTW algorithm to compute distances between the time functions, as in [27]. We modify the consistency value definition in order to make its notation similar to the FDR and thus define the distance discriminant ratio (DDR) R for the ith feature of user u as

$$R_i(u) = \frac{(\mu_{DG_i} - \mu_{DF_i})^2}{\sigma_{DG_i}^2 + \sigma_{DF_i}^2} \qquad (3)$$

where DG_i is the set of distances among the user genuine signatures and DF_i is the set of distances between the genuine signatures and forgeries. This measure assumes that for each user the mean distance between genuine signatures and forgeries μ_{DF_i} is higher than the mean distance between genuine signatures μ_{DG_i}, which has been tested to be true in the datasets used for experiments. As can be seen, while not being mathematically equivalent to the FDR, the DDR provides a comparable measure in terms of the feature discriminative power. Unlike the FDR, this measure is not scale invariant. Consequently, in our experiments local features are normalised to have zero mean and variance equal to 1.

The median FDR and DDR are computed differently for random and skilled forgeries. In the case of random forgeries, for each user, the FDR and DDR between the user samples and the rest of the genuine signatures in the database are computed, whereas for skilled forgeries, the FDR and DDR are computed between the genuine signatures and the available skilled forgeries for each user.

4.2.3 Experiment 3 – feature selection: In order to select the best performing feature combinations, feature selection on the global 100-feature set and the local 25-feature set is carried out using the SFFS algorithm [29],

which is set to minimise the system equal error rate (EER) over the BMDB-DEV50 development dataset.

4.2.4 Experiment 4 – validation: Finally, the verification performance in terms of the EER using the optimal feature vectors selected by the SFFS algorithm for each scenario are compared on the two available validation sets (BMDB-VAL70 and SG-NOTE).

5 Results

5.1 Experiments 1 and 2: individual feature analysis

From Fig. 4*a*, we observe that the median FDR for each feature is similar in the pen tablet and the PDA scenario when random forgeries are considered (left column). Nevertheless, the FDR for pen tablet tends to be always

higher or equal than the FDR for PDA. In the case of skilled forgeries, the FDR is higher in most cases for pen tablet than PDA in the case of skilled forgeries (right column). 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. Interestingly, the FDR for the interpolated pen-ups tablet subset is in general lower than the original subset, especially for skilled forgeries. This suggests that pen-up trajectories are more resilient to forgeries (i.e. harder to imitate).

The DDR is in general higher for pen tablet than for PDA, independently of the availability of pen-up trajectories (see Fig. 4*b*). As for global features, when pen-up trajectories are interpolated, the DDR is more negatively affected for skilled forgeries than for random forgeries. In random forgeries, the most relevant difference is observed in the vertical coordinate feature *y*, which is the one that best
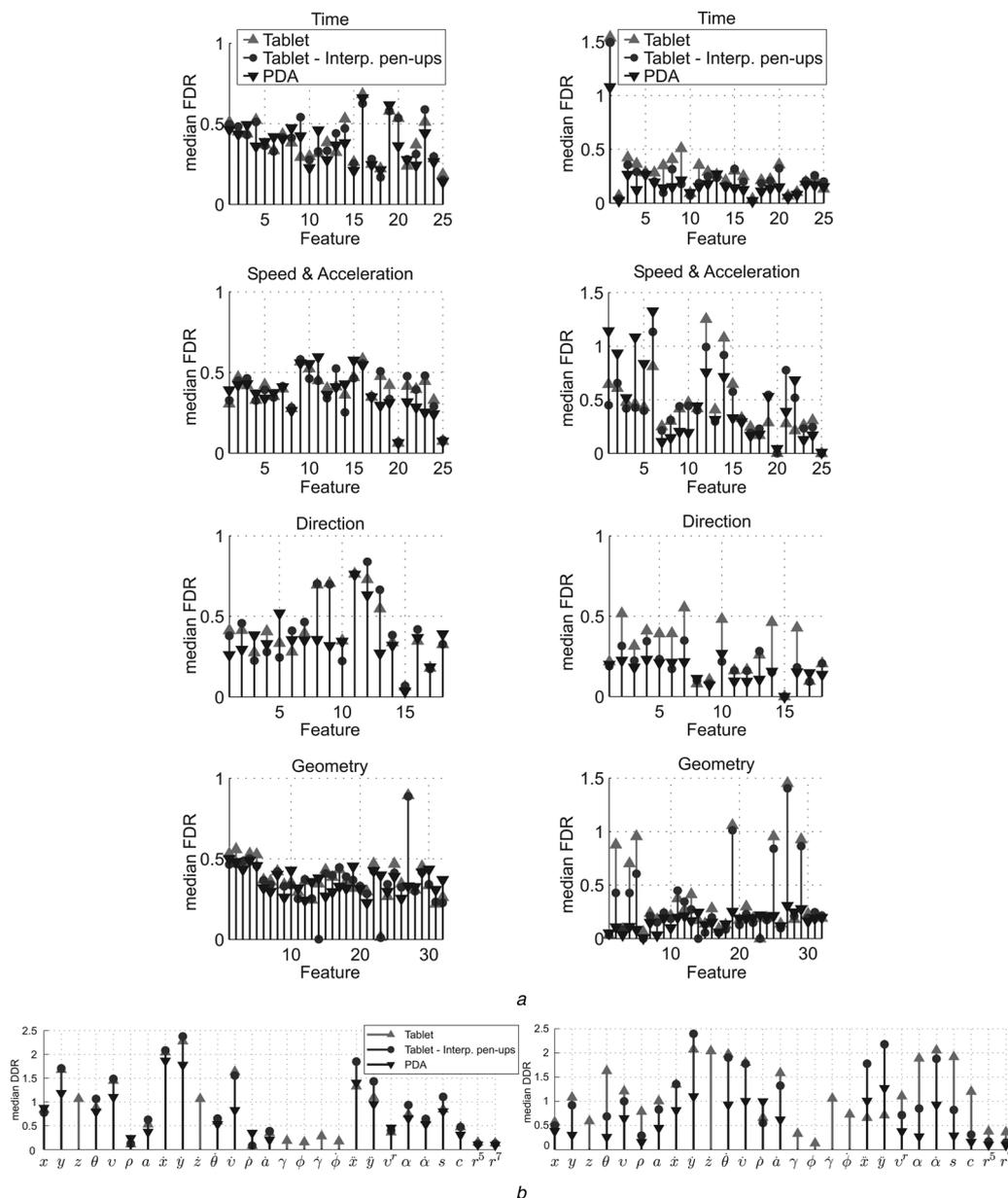
Fig. 4 *FDR and DDR of global and local features*
a FDR of each global feature for random (left) and skilled (right) forgeries
b DDR of each local feature for random (left) and skilled (right) forgeries

characterises the shape of signatures. The first derivative of y has also a notably higher DDR in the pen tablet scenario. This suggests a higher geometrical variability in the PDA scenario. As can be seen, first and second x, y derivatives are more discriminative when pen-ups are interpolated, which may reflect unstable motion during pen-ups. The path velocity magnitude v and its first derivative are also considerably more discriminative in the pen tablet dataset. This suggests higher variability in the writing speed on the PDA, which can be motivated by the unfamiliar signing surface (touch screen) and device.

5.2 Experiments 3 and 4: feature selection

In Fig. 5 the evolution of the global and the local system EER using the optimal feature vector, as selected by the SFFS algorithm, is depicted for each possible vector size. It can be observed that while the behaviour for the case of random forgeries is similar on both scenarios (mobile and tablet), the optimal verification performance is significantly better for skilled forgeries in the pen tablet scenario.

In the global system, the verification performance for pen tablet does not significantly vary when pen-up trajectories are interpolated. On the other hand, the EER increases notably in the local system when pen-ups are interpolated. This corroborates the results from the individual feature analysis, that is trajectories during pen-ups provide considerable discriminative information against skilled forgeries.

5.2.1 Experiment 3 – global features:
As can be seen in Fig. 5a, the optimal feature vectors have an approximate size of 40 features. The specific features which conform the optimal 40-feature vectors are shown in Table 1. The proportion of each feature type (time, speed and acceleration, direction and geometry, as described in Table 1) in each optimisation scenario is represented in Fig. 6, considering feature vectors of 40 elements. As can be seen, geometry features have a higher relevance in the PDA dataset. In contrast, time and speed and acceleration features are more relevant in pen-tablet feature vectors, specially against skilled forgeries. Geometry features are in principle the easiest to forge, so their higher presence in

PDA feature vectors may lead to a lower verification performance.

5.2.2 Experiment 4 – local features:
The optimal local feature combinations selected by the SFFS algorithm for each optimisation scenario are the following:

- *PDA skilled forgeries*: $x, y, v, \rho, \dot{y}, \dot{\rho}, c$
- *PDA random forgeries*: $x, y, \rho, \dot{x}, \dot{y}, \theta, \dot{\alpha}, c$
- *Pen tablet skilled forgeries*: $x, y, v, \dot{y}, \theta, \dot{v}, v^r, \dot{\alpha}, c$
- *Pen tablet random forgeries*: $x, v, \dot{y}, \theta, \dot{\alpha}, c$
- *Pen tablet skilled forgeries interpolated pen-ups*: $x, y, \theta, v, \ddot{y}, v^r, s$
- *Pen tablet random forgeries interpolated pen-ups*: $x, y, a, \dot{y}, v^r, \dot{\alpha}, c$

Several remarks can be extracted from these results. First, neither pressure nor pen orientation-related features are present in the pen tablet optimal feature vectors, suggesting that the lack of them should not penalise the verification performance (in contrast to the results presented in [30]). For the two original datasets (PDA and pen tablet), three features are present in all vectors, namely the x coordinate, the first derivative of the y coordinate and the cosine c of the trajectory angle α.

These results also reveal that less features are needed for HMM-based signature verification compared to the ones commonly considered in other works such as [10, 23, 28], at least under these experimental conditions. The absence of pressure in the optimal feature vectors suggests that a pen tablet-based system does not have a priori advantage over a handheld device because of the capture of pressure information *per se*. The main disadvantage of a handheld device would be the lack of trajectories during pen-ups, which penalises verification performance.

5.3 Experiment 4: validation

The verification performance (in terms of EER) on the BMDB-VAL70 validation set using the optimal feature vectors in each scenario is shown in Table 3. As can be seen, global features provide better results in general on mobile conditions, at least compared to an HMM-based
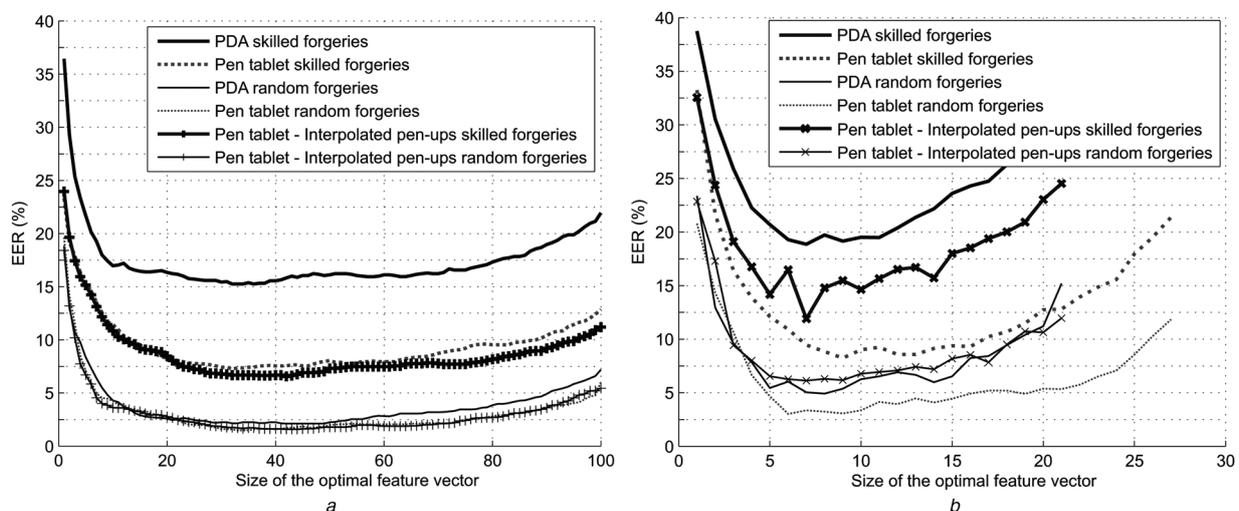
Fig. 5 *System EER for each possible size of the optimal feature vector as selected by the SFFS algorithm for the global and local systems*
a Global system
b Local system
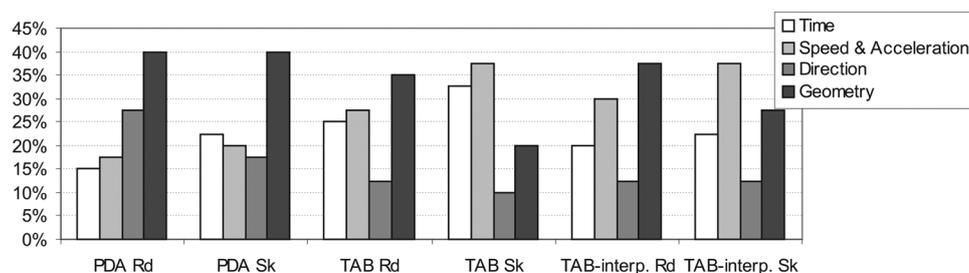
Fig. 6 *Histogram of global feature types selected by the SFFS algorithm in each optimisation scenario*
Feature vectors of 40 elements are considered. Rd denotes random forgeries, Sk skilled forgeries and 'interp.' refers to the interpolated pen-ups dataset

system. It can also be observed that when pen-up trajectories are not available, the performance of the local system is significantly degraded against skilled forgeries. This corroborates the reduction of the individual feature discriminative power (FDR and DDR) against skilled forgeries observed in the individual feature analysis (Section 5.1).

It can also be observed in Table 3 that, comparing both optimisation scenarios, when the systems are optimised against random forgeries, there is a significant degradation in the performance against skilled forgeries. In contrast, the EER against random forgeries is nearly not degraded (or even enhanced) when the systems are optimised against skilled forgeries.

A combined EER (EER_c) is also presented in Table 3, where all available scores (genuine, random forgeries and skilled forgeries) are used for its computation. This implies that, for each user, 15 genuine user scores, 20 skilled forgery scores and 69 random forgery scores are used for the (EER_c) computation. It can be observed that in most cases the systems optimised against skilled forgeries present a better overall performance under this experimental conditions.

In Table 4, the verification performance in terms of EER against random forgeries is shown for the SG-NOTE validation dataset. As can be seen, the performance is similar than in the BMDB database when the local system is used. In contrast, the global system verification performance is better than with the BMDB database.

Results of the BSEC 2009 Signature Evaluation Campaign [12] Task 1 are reported in Table 5. Performance in terms of EER of the eleven participating systems and a reference system is shown. The BMDB signature corpus was used for the competition, which contains 382 users. As can be seen performance is degraded on mobile conditions. The UAM-GLO system is based on the global system presented

Table 4 System performance in terms of EER on the SG-NOTE set using global or local features on both scenarios for random (rd) forgeries

Optimisation scenario		Global EER_{rd}, %	Local EER_{rd}, %
Sk. forgeries	PDA	4.2	6.2
Rd. forgeries	PDA	2.1	6.8

Vectors of 40 features have been selected in every configuration for the global system

Table 5 System performance in terms of EER in the BSEC 2009 Signature Evaluation Campaign both for random (rd) and skilled (sk) forgeries

System ID	DS2 pen tablet dataset		DS3 PDA dataset	
	EER_{sk}, %	EER_{rd}, %	EER_{sk}, %	EER_{rd}, %
UPM1	4.9	2.3	7.4	1.9
UPM2	4.4	1.9	8.2	2.0
SKU	2.9	1.6	7.9	1.3
ASU	3.8	2.7	31.6	30.6
VDU	**2.2**	1.0	6.6	1.7
SU	3.0	2.2	5.0	4.3
UAM-DTWr	4.2	0.5	12.2	0.6
UAM-DTWs	2.9	1.5	5.8	1.5
UAM-HMM	19.2	24.2	25.8	21.3
UAM-GLO	6.7	3.3	13.2	4.7
UAM-FUS	2.2	0.6	5.5	0.7
reference	4.5	1.7	11.3	4.8

Table data have been extracted from [12]

in this work, and the UAM-HMM system is based on the local system. Unfortunately, the UAM-HMM system had an

Table 3 System performance in terms of EER on the BMDB-VAL70 validation set using global or local features on both scenarios for random (rd) and skilled (sk) forgeries

Optimisation scenario		Global			Local		
		EER_{rd}, %	EER_{sk}, %	EER_c, %	EER_{rd}, %	EER_{sk}, %	EER_c, %
Sk. forgeries	PDA	7.2	16.3	9.7	6.0	17.5	9.1
	pen tablet	5.6	11.3	7.5	4.5	9.3	5.7
	pen tablet interp.	6.9	10.9	7.9	6.8	12.1	8.1
Rd. forgeries	PDA	5.4	17.7	9.2	5.8	22.2	9.5
	pen tablet	6.7	13.0	8.6	3.8	11.1	7.1
	pen tablet interp.	6.7	10.9	7.7	5.8	15.3	8.9

Combined EER (EER_c) is also presented, as described in Section 5.3. Vectors of 40 features have been selected in every configuration for the global system

implementation error that led to a poor performance in the BSEC 2009 competition.

6 Conclusions and future work

The effects on the feature discriminative power produced by the usage of handheld devices for signature acquisition have been studied. It has been observed that mobile conditions negatively affect feature discriminative power, specially when local features are considered, at least for the HMM-based system used in the experiments, which is based on the one that reached top positions in the SVC-2004 competition [8].

The performance difference against skilled forgeries between the mobile and pen tablet BMDB datasets may also be because of the different forgery acquisition protocols. On the mobile scenario, forgers had access to an on-screen replay of the signature whereas the replay shown was on a separate screen when using pen-tablet. Nevertheless, it has been clearly seen that verification performance decreases when pen-up samples are not available, except for the case of the global system and skilled forgeries. These results indicate that trajectories during pen-ups contain relevant biometric information, corroborating the findings reported in [31] in the field of handwriting recognition. The verification performance when using global features presents a more robust behaviour than the local approach based on discrete-time functions against the lack of pen-up samples.

It has also been observed that the optimal feature set selected by the SFFS algorithm has a similar performance on the SG-NOTE database in the case of local features, whereas it presents lower error rates for the global system. The performance of the global system is better whereas the local system has a similar performance. This corroborates the apparent robustness of global features against degraded signature acquisition conditions, at least in this experimental setup.

At an individual feature level, it has also been observed that on handheld devices the feature discriminative power is more negatively affected for skilled forgeries than for random forgeries. The discriminative power on the mobile scenario is penalised by the lack of pen-up trajectories, the unfamiliar screen surface where users must sign and the poor ergonomics of a handheld device stylus. Features related to pen inclination and pen pressure, not available in this scenario, have not proven to be among the most discriminant in the pen tablet setup.

An interesting topic for future work is the study of interoperability between devices and the effects of mobile conditions on DTW-based systems, which in recent years have gained popularity in the literature.

7 Acknowledgments

8 References

1 Galbally, J., Martinez-Diaz, M., Fierrez, J.: 'Aging in biometrics: an experimental analysis on on-line signature', *PLoS ONE*, 2013, **8**, (7), p. e69897

2 Impedovo, D., Pirlo, G.: 'Automatic signature verification: the state of the art', *IEEE Trans. Syst. Man Cybern. C, Appl. Rev.*, 2008, **38**, (5), pp. 609–635

3 Plamondon, R., Lorette, G.: 'Automatic signature verification and writer identification: the state of the art', *Pattern Recognit.*, 1989, **22**, (2), pp. 107–131

4 Lee, L.L., Berger, T., Aviczer, E.: 'Reliable on-line human signature verification systems', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1996, **18**, (6), pp. 643–647

5 Fierrez, J., Ortega-Garcia, J.: 'On-line signature verification', in Jain, A. K., Ross, A., Flynn, P. (Eds.): 'Handbook of biometrics' (Springer, 2008), pp. 189–209

6 Nanni, L., Lumini, A.: 'Ensemble of Parzen Window classifiers for on-line signature verification', *Neurocomputing*, 2005, **68**, pp. 217–224

7 Martinez-Diaz, M., Fierrez, J., Ortega-Garcia, J.: 'Universal background models for dynamic signature verification'. Proc. IEEE Conf. on Biometrics: Theory, Applications and Systems, BTAS, 2007, pp. 1–6

8 Kholmatov, A., Yanikoglu, B.: 'Identity authentication using improved online signature verification method', *Pattern Recognit. Lett.*, 2005, **26**, (15), pp. 2400–2408

9 Faundez-Zanuy, M.: 'On-line signature recognition based on VQ-DTW', *Pattern Recognit.*, 2007, **40**, (3), pp. 981–992

10 Fierrez, J., Ramos-Castro, D., Ortega-Garcia, J., Gonzalez-Rodriguez, J.: 'HMM-based on-line signature verification: feature extraction and signature modeling', *Pattern Recognit. Lett.*, 2007, **28**, (16), pp. 2325–2334

11 Vivaracho-Pascual, C., Pascual-Gaspar, J.M.: 'On the use of mobile phones and biometrics for accessing restricted web services', *IEEE Trans. Syst. Man Cybern. C, Appl. Rev.*, 2012, **42**, (2), pp. 213–222

12 Houmani, N., Mayoue, A., Garcia-Salicetti, S., *et al.*: 'BioSecure signature evaluation campaign (BSEC'2009): evaluating online signature algorithms depending on the quality of signatures', *Pattern Recognit.*, 2012, **45**, (3), pp. 993–1003

13 Elliot, S.: 'Differentiation of signature traits vis-a-vis mobile- and table-based digitizers', *ETRI J.*, 2004, **26**, (6), pp. 641–646

14 Fierrez-Aguilar, J., Nanni, L., Lopez-Penalba, J., Ortega-Garcia, J., Maltoni, D.: 'An on-line signature verification system based on fusion of local and global information'. Proc. IAPR Int. Conf. on Audio- and Video-Based Biometric Person Authentication, AVBPA, 2005 (*LNCS* **3546**), pp. 523–532

15 Ortega-Garcia, J., Fierrez, J., Alonso-Fernandez, F., *et al.*: 'The multi-scenario multi-environment Biosecure Multimodal Database (BMDB)', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2010, **32**, (6), pp. 1097–1111

16 Martinez-Diaz, M., Fierrez, J.: 'Signature databases and evaluation', in Li, S.Z. (Ed.): 'Encyclopedia of biometrics' (Springer, 2009), pp. 1178–1184

17 Impedovo, D., Pirlo, G., Plamondon, R.: 'Handwritten signature verification: new advancements and open issues'. Proc. Int. Conf. on Frontiers in Handwriting Recognition, ICFHR, 2012, pp. 367–372

18 TELECOM & Management SudParis, BioSecure Multimodal Evaluation Campaign 2007 Mobile scenario – experimental results, Technical Report, (http://biometrics.it-sudparis.eu/BMEC2007/files/Results_mobile.pdf), 2007

19 Yeung, D.Y., Chang, H., Xiong, Y., *et al.*: 'SVC2004: first international signature verification competition'. Proc. Int. Conf. on Biometric Authentication, ICBA, 2004 (*LNCS* **3072**), pp. 16–22

20 Houmani, N., Garcia-Salicetti, S., Dorizzi, B.: 'A novel personal entropy measure confronted with online signature verification systems' performance'. Proc. IEEE Int. Conf. on Biometrics: Theory, Applications and Systems, BTAS, 2008, pp. 1–6

21 Alonso-Fernandez, F., Fierrez-Aguilar, J., Ortega-Garcia, J.: 'Sensor interoperability and fusion in signature verification: a case study using Tablet PC'. Proc. Int. Workshop on Biometric Recognition Systems, IWBRS, 2005, (*LNCS* **3781**), pp. 180–187

22 Simsons, D., Spencer, R., Auer, S.: 'The effects of constraining signatures', *J. Am. Soc. Questioned Doc. Examiners*, 2011, **14**, (1), pp. 39–50

23 Richiardi, J., Ketabdar, H., Drygajlo, A.: 'Local and global feature selection for on-line signature verification'. Proc. IAPR Eighth Int. Conf. on Document Analysis and Recognition, ICDAR, 2005, pp. 625–629

24 Nelson, W., Kishon, E.: 'Use of dynamic features for signature verification'. Proc. IEEE Int. Conf. on Systems, Man, and Cybernetics, 1991, vol. 1, pp. 201–205

25 Nelson, W., Turin, W., Hastie, T.: 'Statistical methods for on-line signature verification', *Int. J. Pattern Recognit. Artif. Intell.*, 1994, **8**, (3), pp. 749–770

26 Jain, A.K., Nandakumar, K., Ross, A.: 'Score normalization in multimodal biometric systems', *Pattern Recognit.*, 2005, **38**, (12), pp. 2270–2285

27 Lei, H., Govindaraju, V.: 'A comparative study on the consistency of features in on-line signature verification', *Pattern Recognit. Lett.*, 2005, **26**, (15), pp. 2483–2489

28 Van, B.L., Garcia-Salicetti, S., Dorizzi, B.: 'On using the Viterbi path along with HMM likelihood information for online signature verification', *IEEE Trans. Syst. Man Cybern. Part B: Cybernetics*, 2007, **37**, (5), pp. 1237–1247

29 Jain, A.K., Zongker, D.: 'Feature selection: evaluation, application, and small sample performance', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1997, **19**, (2), pp. 153–158

30 Muramatsu, D., Matsumoto, T.: 'Effectiveness of pen pressure, azimuth, and altitude features for online signature verification'. Proc. IAPR Int. Conf. on Biometrics, ICB, 2007 (*LNCS* **4642**), pp. 503–512

31 Sesa-Nogueras, E., Faundez-Zanuy, M., Mekyska, J.: 'An information analysis of in-air and on-surface trajectories in online handwriting', *Cogn. Comput.*, 2012, **4**, (2), pp. 195–205