2010 12th International Conference on Frontiers in Handwriting Recognition

A Comparative Evaluation of Finger-Drawn Graphical Password Verification Methods

M. Martinez-Diaz, C. Martin-Diaz, J. Galbally and J. Fierrez
Biometric Recognition Group - ATVS
Universidad Autonoma de Madrid
Madrid, Spain
marcos.martinez@uam.es, javier.galbally@uam.es, julian.fierrez@uam.es

***Abstract*—Doodle-based graphical passwords represent a challenging scenario due to their high variability and the tendency to be graphically simple. Despite this, doodle-based authentication using touchscreens is a promising lightweight user verification method. Several works have been published in this field, although they report in general experimental verification results over small and private databases. In this paper we analyze the performance of several state-of-the-art systems for doodle verification, using the recently acquired DooDB database, which is publicly available. Several algorithms are tested, from the fields of gesture recognition and doodle and signature verification. A comparative study of their performance is done, and future research directions are pointed out.**

***Index Terms*—Graphical password; doodle; pseudo-signature; passdoodle; touchscreen;**

I. Introduction

Graphical passwords represent a convenient and intuitive user authentication method. Among them, doodle-based graphical passwords have been proposed as an alternative to traditional passwords in touchscreen-enabled devices [1]. Users are validated by tracing a doodle over a touchscreen, which is then accepted or rejected by the system. Due to their graphical nature, they are in general easier to remember than strings composed of characters and numbers [2]. Compared to other user verification methods such as biometrics, doodle-based authentication has some advantages. It only requires a touchscreen, which is now popular in handheld devices, opposed to specific acquisition hardware needed to capture biometric traits such as fingerprints. Within biometrics, signature verification is the most similar trait with respect to doodles. Signature verification on handheld devices has also been studied [3]. As an example of the recent interest in this field, an evaluation campaign with signatures captured on a PDA was organized in 2009, with the participation of several research institutions [4].

Doodle verification shares with signature verification that behavioral information (e.g. dynamics) is used for matching. On the other hand, doodles are commonly invented and thus not composed of natural and trained movements that users have performed for several years. While this may be source of an increased variability, it is also and advantage for doodles since, unlike signatures, they can be easily replaced if necessary (which is known as revocability).

In this paper we evaluate several verification algorithms for doodle verification. These algorithms are selected from the state of the art in gesture recognition, doodle and signature verification. We evaluate if the recent advances in signature verification are also applicable to the problem of doodle-based authentication. The two main approaches for dynamic signature verification are followed. These are global and local systems [5]. Global or feature-based systems model the signature as a holistic vector composed of global features (e.g. average speed, number of pen-ups). Local or function-based systems extract time functions from the signature trajectory (pen coordinates, pressure, etc.) and perform signature matching via elastic or statistical techniques like Dynamic Time Warping (DTW) or Hidden Markov Models (HMM).

Another objective of this work is to obtain a baseline doodle verification performance that can be used to compare this method with other authentication alternatives such as signatures. The recently captured DooDB database is used for experiments. This database is publicly available at the ATVS website [6] and contains doodles and pseudo-signatures (which are finger-drawn simplified signatures) from 100 donors. Some examples of doodles and pseudo-signatures from the database are shown in Fig. 1. We also analyze the differences in the verification performance between doodles and pseudo-signatures. Pseudo-signatures are simplified signatures traced with the finger. They usually consist on some initials or flourish. Since pseudo-signatures are based on real signatures and thus composed of learned movements, it can be hypothesized that they present a lower variability and a better verification performance. The effects of inter-session variability are also studied.

The paper is structured as follows. In Sect. II related works are summarized. The verification systems that are analyzed are described in Sect. III. Experiments are reported in Sect. IV and conclusions are finally drawn in Sect. V.

II. Related Works

Several approaches have been proposed for the problem of doodle-based user authentication. One of the first contributions is the Draw-A-Secret system (DAS) proposed by Jermyn *et al.* [1]. The DAS system implements a grid where users trace their graphical password. The sequence of grid cells that the users follow is then stored and used for validation. Users are validated only if they follow the same sequence of cells. Although the use of a grid with reasonably sized cells

978-0-7695-4221-8/10 $26.00 © 2010 IEEE
DOI 10.1109/ICFHR.2010.65

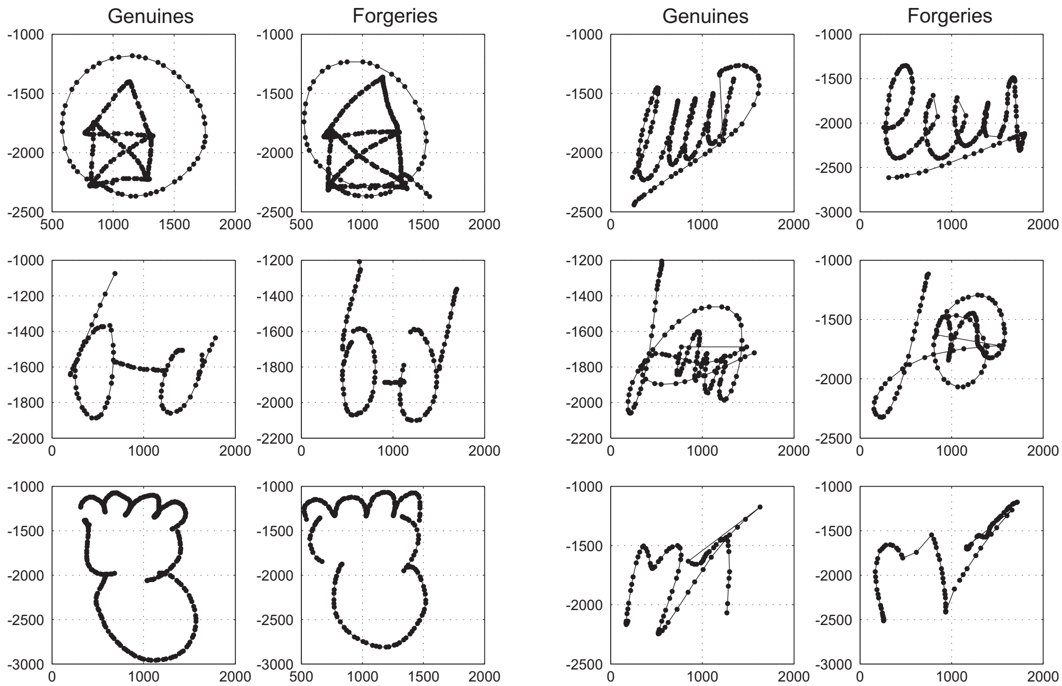

Fig. 1. Example of doodles (left) and pseudo-signatures (right).

allows some variability, the system must deal with problematic passwords with strokes too near to the cell edges. Later, the term "passdoodle" was introduced in [7]. In that work, the memorability of doodles for user authentication is studied. However, it is a preliminary study carried out with doodles traced on a sheet of paper. A passdoodle verification system is proposed in [8]. The stroke spatial distribution and the speed are used for verification. Experiments are performed with a small database containing doodles from 10 users.

In [9], Govindarajulu *et al.* propose a doodle authentication system which uses DTW for matching. Verification performance results are provided using Tamil characters, not real doodles. The Scribble-A-Secret (SAS) scheme for doodle verification was later proposed by Oka *et al.* [10]. This system uses edge orientation patterns as features. Experiments are carried out with doodles from 87 individuals, traced on a Tablet-PC touchscreen.

Chen *et al.* present in [11] a user verification scheme based on predefined visual cues that are chosen by the user with the aid of a graphical interface. With these cues (which are in general common shapes), each user creates what they call a pseudo-signature. Cryptographic keys are then generated from the pseudo-signatures.

In all the aforementioned works, intentional forgeries are not taken into account. Only random forgeries are considered, which is the case where a user claims to be another one but provides his own doodle. Thus, doodles from different users are considered random forgeries of each other.

III. Verification systems under study

Global and local systems for doodle verification are analyzed in this work. Verification systems from the state of the art in signature verification are considered, as well as systems that have been proposed for gesture and doodle recognition.

In all systems doodles are normalized, so their $[x, y]$ sequences have zero mean and variance equal to 1.

A. Global systems

Global feature sets from two works are considered:

- The GRANDMA set, as defined in [12]. This popular feature set was originally proposed for gesture recognition.
- The 100-feature set described in [13] for signature verification. That feature set contains global features selected from several previous works. A 40-feature subset from the 100 features was also considered, which was the one that provided a higher class separability for signatures (as described in [13]). We will refer to these sets as the ATVS-100 and the ATVS-40 in this paper.

Two different distance measures between feature sets are studied, by computing match scores as the inverse of the Euclidean and Mahalanobis distances.

B. Local systems

Two local feature sets are studied. First, the one from the doodle authentication system proposed in [9]. In that system, 6 local features are extracted from the doodle trajectory. These are the coordinate sequence $[x, y]$, and its first and second derivatives (speed and acceleration). Thus, each doodle is described by the 6-dimensional sequence $[x, y, x', y', x'', y'']$.

TABLE I
LOCAL ATVS-BSEC FEATURE SET.

#	Feature	Description
1	x-coordinate	x
2	Second-order derivative of x-coordinate	x''
3	First-order derivative of y-coordinate	y'
4	Second-order derivative of y-coordinate	y''
5	Path velocity	$\upsilon = \sqrt{y' + x'}$
6	First-order derivative of path velocity	υ'
7	First-order derivative of Log curvature radius	ρ', where $\rho = \log(\upsilon/\theta')$ and $\theta = \arctan(y'/x')$ is the curvature of the position trajectory

Matching is performed using the DTW algorithm. We refer to this feature set as HP-LOCAL.

The other system is based on the one presented by the Biometric Recognition Group to the BioSecure Signature Evaluation Campaign BSEC 2009 [4]. In particular, the system is the one based on DTW that was tuned for skilled forgeries. It was one of the best performing in most evaluation scenarios against skilled forgeries. The system extracts 7 local features, which are described in Table I. This feature set is referred to as ATVS-BSEC.

The match score is obtained by averaging the DTW distances between a test sample and the reference set and computing the inverse value. We also study user-dependent score normalization in both systems. For each user, the minimum and average DTW distances, d_{min} and d_{avg} respectively, between the reference samples is computed. In the matching step, the averaged DTW distance between the reference samples and the test sample is divided by d_{avg} for the case of doodles and by d_{min} for pseudo-signatures. This normalization method can be viewed as a intra-user variability compensation technique. The decision of using d_{avg} or d_{min} is taken based on preliminary experimental results which are omitted in this paper for the sake of clarity.

C. Fusion

Fusion of different verification systems is a popular method to increase the verification performance. In this paper, we consider fusion at score level, via a score-weighed sum. Thus, the final score is computed as $s = s_g + k \cdot s_l$, where s_g and s_l are the global and local system scores respectively and k is the fusion weighting factor. The optimal value of k is obtained heuristically during the experiments.

IV. EXPERIMENTS

A. Database and Experimental protocol

The doodle and pseudo-signature subsets from the DooDB database have been used for experiments. The doodle dataset consists of free-form doodles, while the pseudo-signature dataset is composed of simplified finger-drawn signatures. This database consists of samples from 100 users, which have been captured in two sessions separated by an average of 2 weeks. The database was captured in an HTC Touch HD touchscreen-enabled mobile phone at a sampling rate of 100Hz under realistic conditions. Users were requested to hold the handheld device in their own hands while drawing. During each session, each participant provided 15 genuine samples of each type (doodle and pseudo-signature) and 10 forgeries. To increase the quality of forgeries, the system replayed the target sample drawing process on-screen. For each doodle, the $[x, y]$ coordinate sequence is provided, as well as the time interval between each sample. The time interval is in general constant, except in the limits of consecutive strokes.

The first 50 users are selected as the development set, while the rest of users are left for validation of the experimental results. The performance of each configuration is evaluated using the Equal Error Rate (EER).

Enrollment is done with the first 5 genuine samples from the first session of each user. Genuine user scores are computed using the remaining genuine samples. Unless stated otherwise, genuine scores are obtained with the 15 genuine signatures of the second session, to take into account inter-session variability. When the effects of inter-session variability are analyzed, the 10 remaining genuine samples of the first session are used.

Skilled forgery scores are obtained using the 20 available forgeries per user. For each user, random forgery scores are computed by comparing the user reference set or model to one sample of the rest of users. As has been previously stated, random forgeries represent the situation where a forger claims to be another user but uses his own doodle or pseudo-signature.

The effects of interpolation are also studied in the experiments. When the time difference between consecutive samples of a doodle is over 50ms, it is considered that the user starts a new stroke (and raises the finger). Samples during “pen-ups” (using the handwriting term) are linearly interpolated. Results comparing the performance when interpolation is done are also given.

Throughout the next section, when results are presented, EER_{sk} refers to the EER for skilled forgeries and EER_{rd} for random forgeries.

B. Experimental results

1) Global features: First, the performance of the Euclidean distance and the Mahalanobis distance as similarity measures is studied. As can be seen in Table II, the Mahalanobis distance provides better results against skilled forgeries, while the Euclidean distance performs better against random forgeries. The performance of the three global feature sets under study

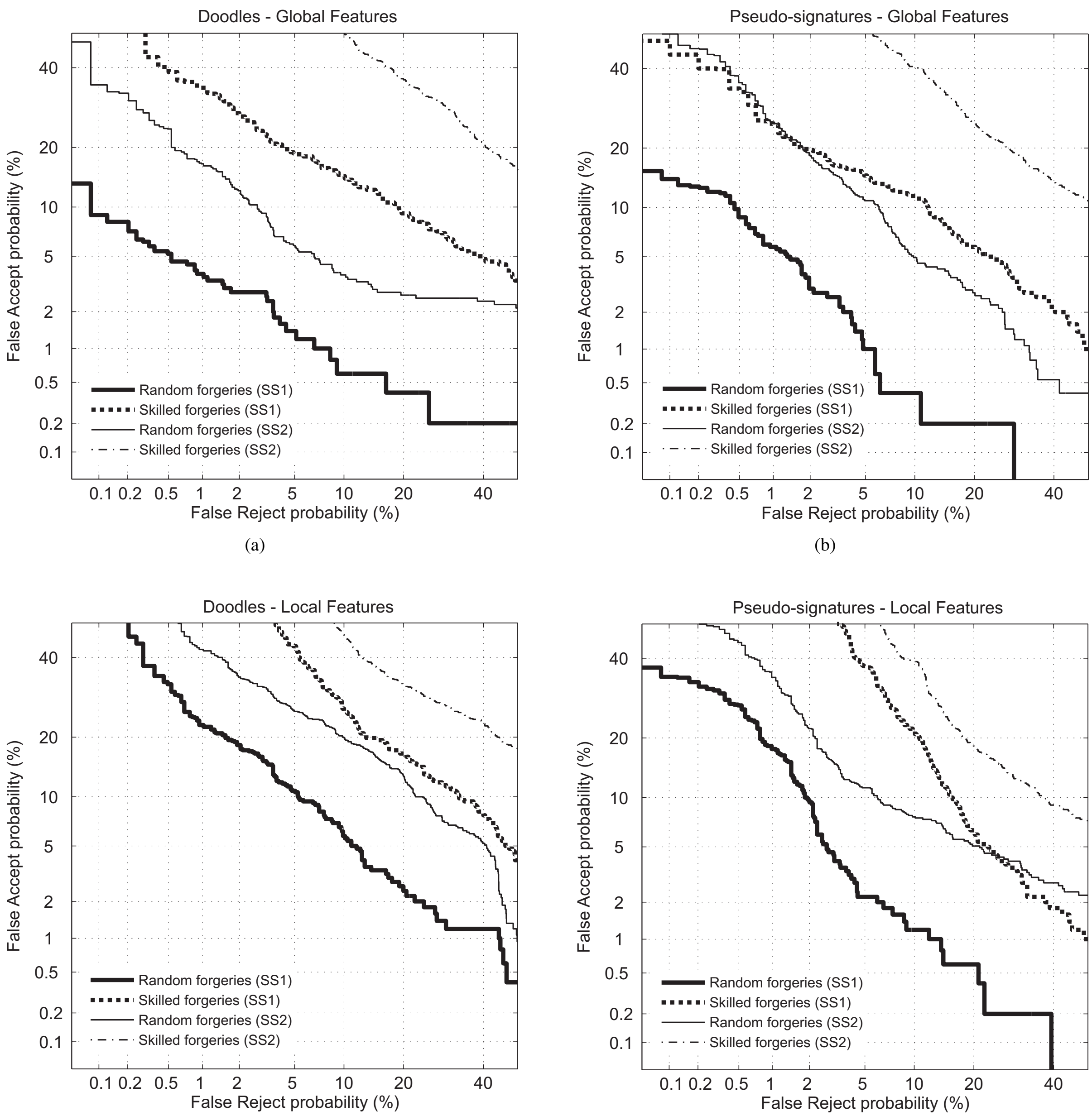

Fig. 2. DET curves using samples from session 1 (SS1) and session 2 (SS2) for testing of the optimal (a) global system for doodles, (b) global system for pseudo-signatures, (c) local system for doodles and (d) local system for pseudo-signatures.

is also shown in Table II. The GRANDMA features, although specifically proposed for gesture recognition are outperformed by the 100- and 40-feature datasets. The best performing set is the 40-feature set, with a notably better performance than the other sets against random forgeries.

Finally, the effect of the proposed interpolation method is studied. As can be seen in Table III, for skilled forgeries, it is slightly beneficial to interpolate samples for pseudo-signatures but not for doodles.

The performance of the best global system (ATVS-40 with Mahalanobis distance) in terms of EER for each modality is shown in Fig. 2 (a) and (b). Results using genuine samples of the first and second session (SS1 and SS2 respectively) are provided). As can be seen, there is a notable performance degradation due to inter-session variability.

2) Local features: In Table IV, the verification performance of the two feature sets considered and the proposed score normalization methods is shown. It is observed that the ATVS-BSEC feature set achieves a better performance for skilled forgeries, although the performance difference is subtle. Re-

TABLE II
VERIFICATION PERFORMANCE FOR DIFFERENT GLOBAL FEATURE SETS AND SIMILARITY MEASURES.

Trait	**Features**	**Euclidean distance**		**Mahalanobis distance**	
		$EER_{rd}(\%)$	$EER_{sk}(\%)$	$EER_{rd}(\%)$	$EER_{sk}(\%)$
Pseudo-signature	GRANDMA	$14,6$	35.0	16.2	27.7
Pseudo-signature	ATVS-100	5.4	28.9	8.8	30.2
Pseudo-signature	ATVS-40	6.6	29.0	**6.7**	**24.1**
Doodle	GRANDMA	12.8	38.3	18.8	29.7
Doodle	ATVS-100	7.8	34.5	8.3	29.5
Doodle	ATVS-40	8.6	34.8	**6.4**	**28.8**

TABLE III
EFFECTS OF THE INTERPOLATION WITH THE ATVS-40 GLOBAL FEATURES.

Trait	EER_{rd}	EER_{sk}
Pseudo-signature, no interpolation	**6.7**	24.1
Pseudo-signature, interpolation	7.3	**22.8**
Doodle, no interpolation	6.4	**28.8**
Doodle, interpolation	**5.0**	29.7

TABLE IV
VERIFICATION PERFORMANCE FOR DIFFERENT LOCAL FEATURE SETS AND DISTANCE MEASURES.

Trait	**Features**	**Traditional DTW**		**DTW & normalization**	
		$EER_{rd}(\%)$	$EER_{sk}(\%)$	$EER_{rd}(\%)$	$EER_{sk}(\%)$
Pseudo-signature	HP-LOCAL	**4.3**	27.0	8.8	23.8
Pseudo-signature	ATVS-BSEC	4.9	26.1	9.6	**23.5**
Doodle	HP-LOCAL	**7.2**	38.9	15.9	33.8
Doodle	ATVS-BSEC	9.7	38.8	16.1	**33.0**

TABLE V
EFFECTS OF THE INTERPOLATION WITH THE ATVS-BSEC LOCAL FEATURES.

Trait	EER_{rd}	EER_{sk}
Pseudo-signature, no interpolation	9.6	23.5
Pseudo-signature, interpolation	**8.4**	**19.0**
Doodle, no interpolation	16.1	33.0
Doodle, interpolation	**15.5**	**27.8**

garding normalization methods, user-dependent score normalization increases the verification accuracy for skilled forgeries. However, it provides significantly worse results for random forgeries.

The effect of the proposed interpolation method is shown in Table V. Unlike for the case of global features, linear interpolation during "pen-ups" provides significant better results in all cases.

The performance of the best performing local systems (ATVS-BSEC with DTW & normalization) is shown in Fig. 2 (c) and (d) for doodles and pseudo-signatures respectively. Results for the first and second session (SS1 and SS2) are also depicted. In general, the performance against random forgeries is worse than with global features. On the other hand, the local systems present a slightly better performance against skilled forgeries.

3) Validation results and fusion: Verification results of the best performing systems against skilled forgeries during experiments for the development set are computed for the validation set (50 last users of the database). As can be seen in Table VI, the variation between development and validation results for the case of doodles is more significant that for pseudo-signatures.

The optimal fusion weighing factor k is computed by selecting the value that provides the minimum EER for skilled forgeries over the development set (50 first users of the database). The value $k = 0.4$ is thus chosen. Verification results of the fusion system (ATVS-40/Mahalanobis + ATVS-BSEC/DTW&normalization) over the validation set are given in Table VII.

V. CONCLUSIONS AND FUTURE WORK

In this paper, doodle verification experiments using several systems based on state-of-the-art approaches of gesture recognition and doodle and signature verification have been performed. These systems are based on global and local features. Results are provided using the recently captured DooDB doodle and pseudo-signature database, which is publicly available.

For our final fusion system, it has been observed that performance of doodles and pseudo-signatures is similar for skilled forgeries, slightly better for pseudo-signatures. The

TABLE VI
VERIFICATION PERFORMANCE OF THE BEST PERFORMING SYSTEMS.

Trait	**Features**	**Development set**		**Validation set**	
		$EER_{rd}(\%)$	$EER_{sk}(\%)$	$EER_{rd}(\%)$	$EER_{sk}(\%)$
Pseudo-signature	Global (ATVS-40)	7.3	22.8	8.9	25.3
Pseudo-signature	Local (ATVS-BSEC)	8.4	19.0	9.7	22.2
Doodle	Global (ATVS-40)	6.4	28.8	4.1	25.8
Doodle	Local (ATVS-BSEC)	15.5	27.8	9.6	24.6

TABLE VII
VERIFICATION PERFORMANCE OF THE FUSION SYSTEM (ATVS-40 + ATVS-BSEC) OVER THE VALIDATION SET.

Trait	**Features**	EER_{rd}	EER_{sk}
Pseudo-signature	Fusion	5.8	20.0
Doodle	Fusion	2.5	22.2

opposite occurs for random forgeries, being in this case much better the doodles. The performance of the individual systems adopted from the signature verification state of the art have reached better results than the gesture-based approaches in all cases. Also interestingly, the usage of local features does not provide a significant better performance with respect to global features, which is usually the case in signature verification. This may be due to a higher variability in the temporal order of the strokes and their dynamics compared to signatures. While signatures are composed of natural movements, doodles are in general invented.

It has been observed that inter-session variability highly degrades the verification performance. This effect may be higher than in signature verification due to the fact that the drawings from the DooDB database are relatively new for the participants.

The verification performance rates that have been obtained have room for improvement, but represent promising values given the simplicity of doodles. The worst-case scenario for forgeries must be taken into account, since the target samples are replayed on-screen.

Future work includes the analysis of variability compensation techniques and the usage of feature selection techniques to find robust features in this challenging scenario.

ACKNOWLEDGMENT

J. G. is supported by a FPU fellowship from Spanish MEC and J. F. is supported by a Marie Curie fellowship from the European Commission. This work was supported by project Bio-Challenge (TEC2009-11186), Contexts (S2009/TIC-1485) and BIO-PASS (TEC2006-13141).

REFERENCES

[1] I. Jermyn, A. Mayer, F. Monrose, M. K. Reiter, and A. D. Rubin, "The design and analysis of graphical passwords," in *Proc. of 8th USENIX Security Symp.*, 1999.

[2] K. Renaud, "On user involvement in production of images used in visual authentication," *Journal of Visual Languages & Computing*, vol. 20, no. 1, pp. 1 – 15, 2009.

[3] M. Martinez-Diaz, J. Fierrez, J. Galbally, and J. Ortega-Garcia, "Towards mobile authentication using dynamic signature verification: useful features and performance evaluation," in *Proc. Intl. Conf. on Pattern Recognition, ICPR*, 2008, pp. 1–6.

[4] TELECOM & Management SudParis. Biosecure Signature Evaluation Campaign, BSEC 2009. [Online]. Available: http://biometrics.it-sudparis.eu/BSEC2009

[5] M. Martinez-Diaz, J. Fierrez, and S. Hangai, *Encyclopedia of Biometrics*. Springer, 2009, ch. Signature features, pp. 1185–1192.

[6] http://atvs.ii.uam.es/.

[7] J. Goldberg, J. Hagman, and V. Sazawal, "Doodling our way to better authentication," in *Proc. of CHI '02 extended abstracts on Human factors in computing systems*, 2002, pp. 868–869.

[8] C. Varenhorst, "Passdoodles; a lightweight authentication method," Research Science Institute, Massachusetts Institute of Technology, Tech. Rep., 2004. [Online]. Available: http://people.csail.mit.edu/emax/papers/varenhorst.pdf

[9] N. S. Govindarajulu and S. Madhvanath, "Password management using doodles," in *Proc. of 9th Intl. Conf. on Multimodal Interfaces, ICMI*, 2007, pp. 236–239.

[10] M. Oka, K. Kato, X. Yingqing, L. Liang, and F. Wen, "Scribble-a-secret: Similarity-based password authentication using sketches," in *Proc. of Intl. Conf. on Pattern Recognition, ICPR*, Dec. 2008, pp. 1–4.

[11] J. Chen, D. Lopresti, L. Ballard, and F. Monrose, "Pseudo-signatures as a biometric," in *Proc. of 2nd Intl. Conf. on Biometrics: Theory, Applications and Systems, BTAS 2008.*, 2008.

[12] D. Rubine, "Specifying gestures by example," *SIGGRAPH Computer Graphics*, vol. 25, no. 4, pp. 329–337, 1991.

[13] J. Fierrez-Aguilar, L. Nanni, J. Lopez-Penalba, J. Ortega-Garcia, and D. Maltoni, "An on-line signature verification system based on fusion of local and global information," in *Proc. of IAPR Intl. Conf. on Audio- and Video-Based Biometric Person Authentication, AVBPA*. Springer LNCS-3546, 2005, pp. 523–532.