Towards human-assisted signature recognition: improving biometric systems through attribute-based recognition

Derlin Morocho
Departamento de Electrica y Electronica, Universidad de las Fuerzas Armadas-ESPE, Sangolquí, Ecuador
dmorocho@espe.edu.ec

Aythami Morales, Julian Fierrez, Ruben Vera-Rodriguez
ATVS- Biometric Recognition Group
Universidad Autonoma de Madrid, Spain
{aythami.morales,julian.fierrez,ruben.vera}@uam.es

Abstract

This work explores human-assisted schemes for improving automatic signature recognition systems. We present a crowdsourcing experiment to establish the human baseline performance for signature recognition tasks and a novel attribute-based semi-automatic signature verification system inspired in FDE analysis. We present different experiments over a public database and a self-developed tool for the manual annotation of signature attributes. The results demonstrate the benefits of attribute-based recognition approaches and encourage to further research in the capabilities of human intervention to improve the performance of automatic signature recognition systems.

1. Introduction

The signature is worldwide accepted as an identity authentication method and it has been used by several cultures over the past 2000 years. The signature is a behavioral biometric trait which comprises neuromotor characteristics of the signer (e.g., our brain and muscles among others define the way we sign) as well as socio-cultural influence (e.g., the Western and Asian styles). During centuries, the examination of signatures has been made by experts who determine the authenticity of the sample based on forensic analysis. More recently, automatic signature verification systems (ASV) emerged as a feasible way to automate the traditional signature verification method made by Forensic Document Examiners (FDEs) [1][2][3].

The variety of applications of automatic signature recognition systems is large. In most of these applications, humans usually supervise the signing process but their responsibilities are mostly limited to guarantee the correct record of the data without any contribution to recognition. These supervisors do not usually have the specific experience of FDEs and they will be refered to as layman in the rest of this work. The deployment of automated systems is eliminating human intervention in many recognition applications. However, perception and analytic capability of humans must not be undervalued and there is large room for improvement compared to fully automatic methods when one applies both the computers and human abilities in some scenarios. Some of these scenarios where a layman may help or contribute to an automatic signature verification are banking, point of sales, notary public, or parcel delivery. We advocate for the consideration of human interaction in these scenarios due to the particularities of the dynamic signature as a behavioral biometric. As it has been demonstrated [3], this biometric fluctuates severely for different users and acquisition conditions. Our aim in this research line of human interactions in automatic systems is to alleviate such fluctuations with simple actions a layman can take in many scenarios of practical importance. Which actions to take and to what extend those actions can help state-of-the-art Automatic Signature Verification systems (ASV) is the final aim of this research line. The layman intervention on ASVs can be done at multiple levels or phases of the biometric system, see Fig. 1. The potential layman interventions include: quality assessment to remove bad quality samples, feature annotation, classification of samples or decision support, among others. However, the performance of laymen in signature verification tasks remains unexplored and their capabilities undervalued [4].

The present work analyzes the potential of attribute-based manual signature recognition and the final aim (not covered in this work) is to obtain a reduced set of high discriminative features which can be either automatically extracted or manually labeled in a reasonable amount of time (e.g., less than 10 seconds for a point of sales or less than 1 minute for an important banking transference). The contributions of this work are threefold: i) the establishment of human baseline performance (layman) in signature recognition tasks based on crowdsourcing experiments; ii) a novel attribute-based signature recognition system via human intervention; and iii) a combined scheme incorporating the proposed attribute-based manual approach to a state-of-the-art automatic online signature recognition system.

The rest of the work is organized as follows: Section 2 summarizes related works. Section 3 presents our work to establish a human baseline performance. Section 4 describes the proposed manual attribute-based signature recognition. Section 5 reports the experiments and results. Finally, Section 6 summarizes the conclusions and future works.

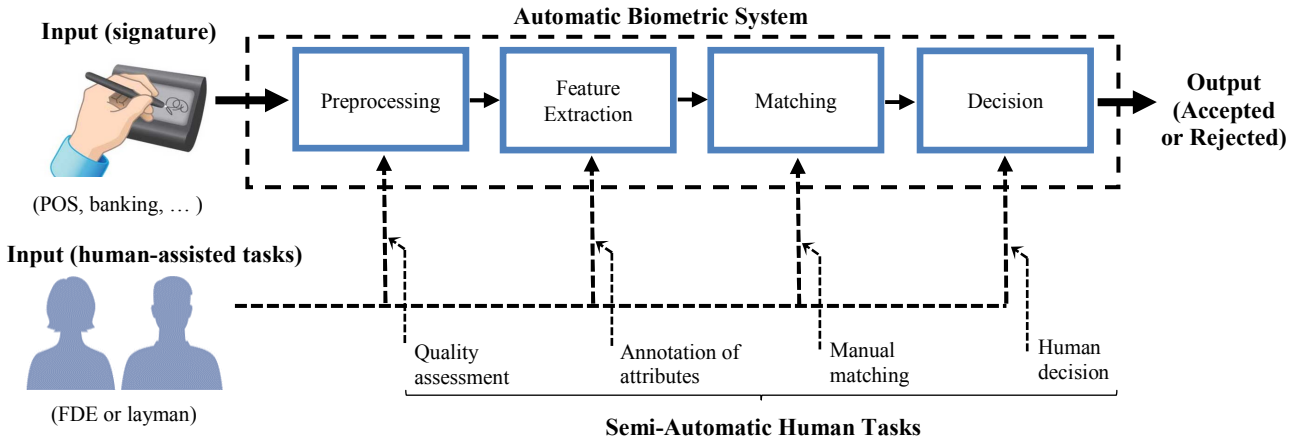

Figure 1. Human-assisted signature recognition basic scheme.

2. Human-assisted signature recognition

Human–assisted schemes in biometrics take advantage of both human abilities and automated system capabilities [5][6][7][8][9]. The use of human annotations in automatic biometric recognition systems has provided encouraging results in the literature [9]. Attribute annotation made by humans has emerged as a way to improve automatic recognition systems in face [5][7][8][9] or gait [6] recognition.

The attribute annotation of signatures is a common task in FDEs analysis and it consists of either discrete labels (e.g.,, the signature has proper punctuation) or scalar measures of specific characteristics (e.g.,, a stroke length of 6 mm). Oliveira et al. [10] analyzed the performance of graphology features in automatic signature verification with promising results over a dataset with 5600 signatures from bank checks. Malik et al. [4] compared the performance of FDE and automatic signature recognition systems for disguised signatures. The results obtained in their study suggest that FDE can achieve similar performance to automatic systems with accuracies over 90%. FDEs are well trained to analyze the authenticity of signatures and their performance is usually high classifying genuine and forged samples [11]. Although the experience of the expert is also exploited, the analysis of FDE is mostly based on well-defined protocols and methodologies. The set of features proposed by FDEs is large [10][12][13] and there is no universal protocol. The results of FDE evaluations are therefore a mix of experience, training and personal subjectivity. It is reasonable to assume that the analysis performed by a non FDE human (once excluded the experience and training) could be centered on the personal subjectivity of each subject. While the baseline performance of FDEs has been analyzed in the literature [14], to the best of our knowledge, the literature lacks of studies analyzing the baseline performance of laymen.

3. Establishing human baseline performance via crowdsourcing

The use of our signature in our day-to-day life make us good forgery detectors of imitation (made by others) of our own signature. We are capable to differentiate our intra-person variability from the variability of a forger (our brain models are trained with hundreds of samples made during years of practice). This ability can be extended to signatures from other people but it is expected a drop of performance caused by the lack of information about the variability of the owner. In [11][15] the ability of 22 individuals was evaluated using 51 signatures (15 samples per individual mixing genuine and forged samples). The results obtained suggest that people perform worse than a state-of-the-art offline signature recognition automatic system (HMM-based Equal Error Rate of approximately 12%).

Here we apply crowdsourcing for the recruitment of amateur volunteers (people without FDE previous experience) to establish a performance baseline of laymen in signature authentication tasks. Amazon Mechanical Turk (MTurk) is a popular web-platform for the acquisition of data through Human Intelligence Tasks (HITs). The participants (workers) provide responses to the requesters. We designed a simple experiment involving 60 workers, and 60 signatures (from 20 different signers) from BiosecurID database [16]. Two samples are showed to the worker: one labeled as genuine signature and the authenticity of the second one remains sequestered (30 genuine and 30 forged signatures are showed in random order). The task consists on labeling the second signature as genuine, forged or not defined (the worker does not know the response). Only the signature image is shown and no instructions on which features to exploit are given. The idea is to analyze the performance of humans (laymen) and their ability to detect forgeries (see Table 1).

Table 1. Human performance obtained from the MTurk task

	FRR	FAR	ND
Human Performance	26.7%	30.0%	33.3%

The results suggest the difficulties of laymen to correctly recognize between genuine and forged signatures. Without any specialized training, many workers doubt when they have to define the authenticity of a given signature (note the high percentage of not defined responses). The poor performance may be alleviated by considering multiple genuine reference signatures (this will be explored in the future).

4. Attribute-based signature recognition

The human baseline performance obtained (see Table 1) suggests that laymen have difficulties to correctly recognize the authenticity of signatures. However, it is well accepted that FDEs can achieve competitive performances based on their specialized training and experience. We propose to analyze selected manual annotated attributes (made by laymen but inspired by the work of FDEs) as features recognition.

An application has been developed (Matlab2012© GUI) to gather several sources of information about the signatures inspired by the work conducted by FDEs. The application is designed to be used by a human without previous experience on FDE analysis or signature recognition tasks. All attributes are annotated from a unique static binary image of the signature (dynamic data is not employed and each signature is annotated separately). The features annotated using the application are described below.

4.1. Signature attributes

There are many features of a signature that can be analyzed [10][12][13]. Due to the large number and variety of existing features, we have selected a set of 13 attributes (inspired in the FDE analysis) on the basis of two characteristics: i) efficiency: the annotation of the attributes must be fast for a layman without any FDE experience; ii) performance: the attributes must be discriminative and useful for signature recognition. We divided the set of features into two groups:

- **Categorical attributes (A1-A9)**: denoted by discrete labels (e.g., spaced/concentrated signature or proportional/unproportioned signature).
- **Scalar Measures (A10-A13):** which are calculated according representative keypoints manually located (e.g., distance between characters or strokes). The keypoint selection reflects the human ability to highlight most representative signature regions.

The set of features allow to explore how the human perception can help to improve automatic systems. Guidelines (in form of a few examples) are shown to the annotator to obtain more consistent features. Listed below are the features chosen, based on the principles given above:

Shape (A1): this attribute is associated with the graphical model used to create the signature (focused on the contour of the signature). The labels associated to this attribute are: rounded strokes, vertical strokes, horizontal strokes, calligraphic model, vertical and horizontal strokes, or unknown.

Proportionality (A2): the proportion is related to the symmetry and size of the handwriting: proportional, unproportioned, mixed or unknown.

Text-loops (A3): predominant style of the loops (typical in letters such as "l, g, p, f, j, y" and others) and directional changes (typical in uppercase letters such as "A, M, N" and others). The possible labels are: round, sharp or unknown.

Order (A4): this attribute refers to the graphic distribution of the parts that form the signature: clear order, confusing, concentrated or spaced.

Punctuation (A5): this attribute analyzes any punctuation mark or distinctive stroke that can characterize the signature (e.g., "i" or "j" punctuation): the signature has proper punctuation, the signature has punctuation but in the wrong place or there is not punctuation.

Flourish-characteristics (A6-A8): we include three attributes related to flourish features. These attributes are symmetry of the most representative loops in the flourish (symmetric, asymmetric and unknown), weight (thin, wide, and unknown) and roundness (round, sharp and unknown).

Hesitation (A9): this attribute reveals the level of perceived hesitation in the signature. Hesitation produces enlargement of characters, tendency of curves to become angles, patching and retouching, tremors, among others. We set three labels for this attribute: the user didn't hesitate while making the signature, the user did hesitate while making the signature or unknown.

Alignment to the baseline (A10): also known as slant it is easy to calculate in some signatures (those with elongated shape) but it could be a challenge in signatures with high complexity (disruptive text and flourish). It is defined as the angle between the main dominant axis of the signature and the baseline.

Slant of the strokes (A11): this attribute measures the slope (angle respect the baseline) of up to three different characters or stroke segments. The annotator has to choose which are the most relevant strokes (if they exist, otherwise the attribute is set to zero).

Strokes-length (A12): as in the slant measures, the annotator has to select up to three representative strokes (initial and ending points) to automatically calculate their lengths (in pixels).

Character spacing (A13): this attribute measures the separation (in pixels) between up to four most relevant characters in the signature.

We are aware that the proposed number of attributes is

large and the most appropriate subset (based either on performance or efficiency) should be chosen once established the potential of this novel recognition technique and the target application (e.g., time and performance requirements vary for different applications).

4.2. Attribute-based matching

All attributes are combined into a unique vector for each signature. The distance between two attribute vectors is calculated using the Manhattan distance normalized by the average absolute deviation of each attribute. A discrete value between 0 and 6 (depending of the number of labels of the attribute) is assigned to each categorical feature. The scalar measures are values which depend on the attribute (length in pixels for A12-A13 or angles in radians for A10-A11). Assume $\mathbf{f} = [f_1, f_2, \dots, f_M]$ as the feature vector (with M features) of a given test sample and $\mathbf{g}^k = [g_1^k, g_2^k, \dots, g_M^k]$ $k \in 1, \dots T$ as an enrollment set with T samples. The distance between the feature vector $\mathbf{f}$ of the test sample and the enrollment set $\{\mathbf{g}^k\}_{k=1}^T$ is calculated as:

$$d = \sum_{i=1}^{M} |f_i - \bar{g}_i|/\sigma_i \qquad (1)$$

where $\bar{\mathbf{g}}$ is the average of the enrollment set and $\boldsymbol{\sigma} = [\sigma_1, \sigma_2, \dots, \sigma_M]$ is the standard deviation of the enrollment features. In our experiments T is equal to 4 and $M = 20$ (note that attributes A11-13 comprise 10 measures).

5. Experiments

The experiments are designed to answer the following questions: What is the performance of manual annotated signature attributes? What is the complementarity (in terms of performance) between human attribute-based recognition and traditional automatic online signature recognition?

5.1. Protocol and baseline performance

The database used in our experiments is a subcorpus of the signature data in the BiosecurID multimodal database [16]. The complete BiosecurID database was acquired in five different universities using five different acquisition devices. To avoid any bias in the results, only the UAM subcorpus, which is the largest subcorpus within the database, will be considered in this work. The subcorpus employed comprises the first 30 signers of the UAM corpus acquired in 4 different sessions, with 16 genuine signatures (four per session) and 12 simulated forgeries (three per session) for every subject (30 × 28 = 840 signatures). Signatures were performed on a marked area over paper templates (25 mm × 120 mm) with an inking pen which also captured the **x** and **y** trajectories and the pen pressure **p** during the signing process, with a sampling frequency of 100 Hz.

Table 2. Performance (% EER) of DTW algorithm over UAM-BiosecurID database

BiosecurID set	EER	
	Random	Simulated
Complete set (132 signers)	0.75%	6.28%
Subcorpus (first 30 signers)	1.93%	6.94%

The 840 signatures were manually annotated according to Sect. 4.1. The annotation was made by an MSc student without any previous experience on FDE analysis. No information about the authenticity (genuine or imitation) of the samples was provided to the annotator and all signatures were analyzed separately (one by one sorted by number of user in the database). The experiment is divided into two categories:

- **Scenario 1 - Random Forgery**: the model of the user is evaluated using genuine samples from other users (different to the owner) as impostor attacks (simulation of users who try to spoof the identity of the user with their own signature).
- **Scenario 2 - Simulated Forgery**: also known as skilled forgeries, the model of the user is evaluated using imitations made by other users (with different level of skill, see the database description for details [16]).

The genuine samples of the first session (4 signatures) are employed as enrollment set and the rest for testing. Hence, the experiments include 12 × 30 = 360 genuine comparisons, 8 (second and third sessions) × 30 × 29 = 6960 random forgery comparisons and 12 × 30 = 360 simulated forgeries comparisons.

In order to compare the performance of attribute-based signature recognition and state-of-the-art online signature recognition systems we have used a function based Dynamic Time Warping algorithm (ranked among top three algorithms in international technology evaluations [17][18]). The algorithm is based on DTW [19] applied to functions of time sequences extracted from each signature. A set of seven time functions are derived from [**x**, **y**, **p**] sequences. The sequences were selected after feature selection (based on the performance of the feature set) from a larger set of sequences defined in [20]. The DTW algorithm matches two different set of sequences based on the Euclidean distance between the time functions (five correspondences among samples of the two sequences are allowed). The classification score is worked out as the average distance between one test signature and the enrolled set. As a baseline, Table 2 shows the performance of DTW algorithm over the complete dataset and the subcorpus employed. Note that the subset of the BiosecurID database employed (first 30 signers) includes some challenging samples which degrade the performance in comparison with the entire database.

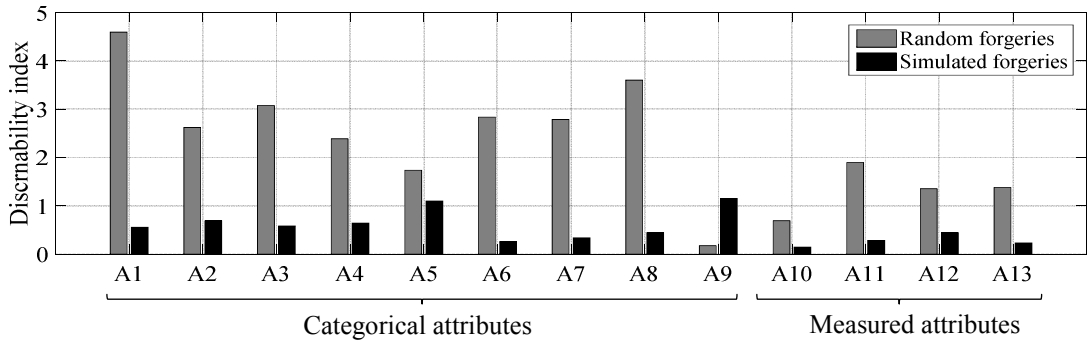

Figure 2. Discriminability index of the different attributes for random and simulated comparisons.

5.2. Experimental Results

The first experiment is carried out to ascertain the discriminative power of manual annotated attributes. The values of the attribute A are first normalized as:

$$A' = \frac{1}{2}\left(\tanh\left(0.01\left(\frac{A-\mu^A}{\sigma^A}\right)\right)+1\right) \quad (2)$$

where μ^A and σ^A are the attribute mean and standard deviation from the genuine signatures. Given $N = 30$ the number of signers, we define two discriminability indexes D_R and D_F for random and simulated forgeries respectively. D_R is computed for a specific attribute A' as:

$$D_R = \frac{1}{(N-1)N}\sum_{i=1,i\neq j}^{N}\sum_{j=1}^{N}\frac{\left|\mu^i-\mu^j\right|}{\sigma^i+\sigma^j} \quad (3)$$

where μ^i is the attribute mean for signer i computed across the 16 available genuine signatures per signer. Similarly σ^i is also the attribute standard deviation for signer i. The discriminability index of simulated forgeries D_F is computed as:

$$D_F = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|\mu^i-\mu_F^i\right|}{\sigma^i+\sigma_F^i} \quad (4)$$

where μ_F^i and σ_F^i are the mean and standard deviation of the simulated forgeries of the signer i computed across the 12 available forgeries per signer. As is expected, the discriminability of attributes is higher in random forgeries, see Fig. 2. However, the results suggest that depending of the scenario (random or simulated forgeries), some attributes can be more discriminant than others. As an example, the Hesitation (A9) is more discriminant for simulated forgeries than for random. This is because of the vacillations of the forger which are not present in genuine signatures (used for the random comparisons). On the other hand, the Shape (A1) is highly discriminative for random forgeries but not for simulated.

The rest of the experiments try to ascertain the performance of the manual annotated signature attributes (detailed in *Section 4*) and the improvement obtained when it is combined with an online signature recognition system (DTW-based). Table 3 shows the performance of the systems and Fig. 3 shows the ROC curves obtained for the different systems and scenarios. Before the combination of systems, the scores obtained by the attribute-based and the DTW online system are normalized using the min/max technique. It allows to normalize the ranges of the classification score to [0-1] values. Finally, the normalized scores are combined according to the sum rule. The results suggest that attribute-based signature recognition is a very promising new technique. Although its performance is clearly lower in comparison with DTW-based recognition, the combination boosts the performance up to 25% (simulated forgery) and 90% (random forgery).

Table 3. Performance (% EER) for the different systems on the BiosecurID subcorpus

System	EER	
	Random	Simulated
Automatic DTW	1.93%	6.94%
Manual Attribute-based	2.46%	13.33%
Combination (DTW+Attribute)	0.06%	5.00%

6. Conclusions and future work

This work explores human-assisted signature recognition including a baseline performance of human recognition of signatures and the analysis of manual attribute-based signature recognition. The results suggest the potential of these recognition schemes in applications involving human interventions. Although the comparison is made using dynamic data, the application of attribute-based signature recognition to offline signature is direct (dynamic data is not used for the annotation of attributes). The performance of offline signature recognition systems is lower than online systems [21] and therefore, attribute-based matching could help to overcome the limitations of offline matching. However, there is large room for further research in this area and the number of open questions is large: What is the consistency of attributes annotated by different users? The consistency between annotations made by different people should be analyzed (experiments involving 20 annotators are currently in course). What are the most discriminant attributes? A smart feature selection can improve the performance of the whole set of attributes. How scalable is

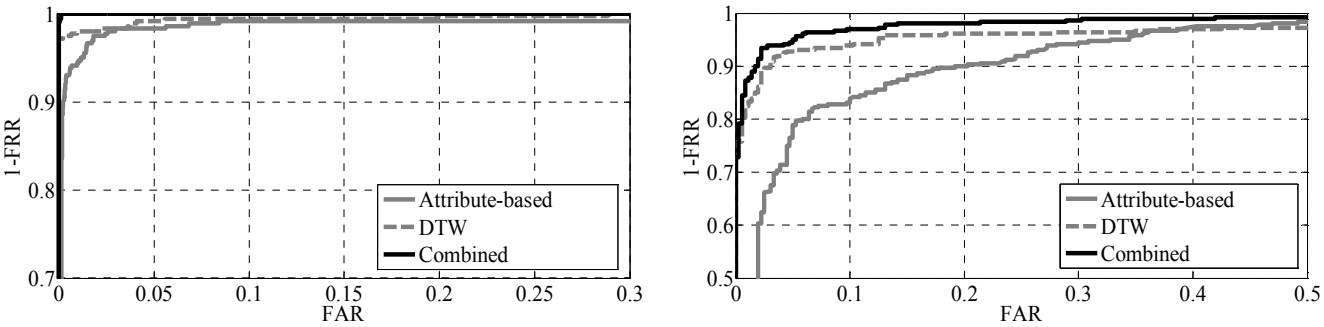

Figure 3. ROC curves for the different forgery scenarios (random forgery on the left and simulated forgery on the right) and systems.

the system when the experiments involve a larger dataset? What is the performance compared with offline signature recognition systems? Why not adding attributes based on the dynamics of the signature (e.g., velocity or pressure profiles)?

Acknowledgment

A.M. is supported by a JdC contract by the Spanish MECD (JCI-2012-12357). This work has been partially supported by projects: Bio-Shield (TEC2012-34881) from Spanish MINECO, BEAT (FP7-SEC-284989) from EU.

References

[1] R. Plamondon and S. N. Srihari. On-line and off-line handwriting recognition: A comprehensive survey. IEEE Trans. on Pattern Analysis and Machine Intelligence, 22:63-84, 2000.

[2] D. Impedovo and G. Pirlo. Automatic signature verification: The state of the art. IEEE Trans. on Systems, Man, and Cybernetics (Part C), 38(5):609-635, 2008.

[3] J. Fierrez and J. Ortega-Garcia. On-line signature verification. A. K. Jain, A. Ross and P.Flynn (Eds.), Handbook of Biometrics, Springer, pp. 189-209, 2008.

[4] M. I. Malik, M. Liwicki, A. Dengel. Part-based automatic system in comparison to human experts for forensic signature verification. Proc. Int. Conf. on Document Analysis and Recognition, Washington DC, USA, pp. 872–876, 2013.

[5] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Describable visual attributes for face verification and image search. IEEE Trans. Pattern Analysis and Machine Intelligence, 33(10):1962–1977, 2011.

[6] D. Reid, M. Nixon and S. V. Stevenage. Soft Biometrics; Human Identification using Comparative Descriptions. IEEE Trans. on Pattern Analysis and Machine Intelligence, 36(6): 1216-1228, 2014.

[7] B. F. Klare et al. Suspect Identification Based on Descriptive Facial Attributes. Proc. of International Joint Conference on Biometrics, Clearwater, Florida, USA, pp. 1-8, 2014.

[8] P. Samangouei, V. M. Patel and R. Chellappa, Attribute-based Continuous User Authentication on Mobile Devices. Proc. Int. Conf. on Biometrics: Theory, Applications and Systems, Washington DC, USA, 1-6, 2015.

[9] P. Tome, J. Fierrez, R. Vera-Rodriguez and M. Nixon. Soft Biometrics and their Application in Person Recognition at a Distance. IEEE Trans. on Information Forensics and Security, 9(3):464-475, 2014.

[10] L. Oliveira, E. Justino, C. Freitas, R. Sabourin. The graphology applied to signature verification. Proc. 12th Conf. of the Int. Graphonomics Society, Salerno, Italy pp. 286-290, 2005.

[11] J. Coetzer, B.M. Herbst, J.A. Du Preez. Off-line signature verification: A comparison between human and machine performance. Proc. 10th Int. Workshop on Frontiers in Handwriting Recognition, La Baule, France, pp. 481-485, 2006.

[12] T. M. Burkes, D. P. Seiger and D. Harrison. Handwriting examination: Meeting the challenges of science and the law. Forensic Science Communications, 11(4), 2009.

[13] E2290-07a Standard Guide for Examination of Handwritten Items, ASTM, 2007.

[14] M. I. Malik, M. Liwicki, A. Dengel, and B. Found. Man vs. Machine: A Comparative Analysis for Forensic Signature Verification. Proc. of the 16th International Graphonomics Society Conference, pp. 9–13, 2013.

[15] H. Coetzer and R. Sabourin. A human-centric off-line signature verification system. Proc. Int. Conf. on Document Analysis and Recognition, Curitiba, Brazil, pp. 153-157, 2007.

[16] J. Fierrez, et al. BiosecurID: a multimodal biometric database. Pattern Analysis and Applications, 13(2):235-246, 2010.

[17] N. Houmani, A. Mayoue, et al., Biosecure signature evaluation campaign (BSEC2009): evaluating online signature algorithms depending on the quality of signatures, Pattern Recognition, 45:993–1003, 2012.

[18] M. I. Malik, et al. ICDAR2013 competitions on signature verification and writer identification for on- and offline skilled forgeries (SigWiComp2013). Proc. of Int. Conf. on Document Analysis and Recognition, Tunisia, pp.1108-1114, 2013.

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

[20] M. Martinez-Diaz, J. Fierrez, R.P. Krish and J. Galbally. Mobile signature verification: feature robustness and performance comparison. IET Biometrics, 3:267–277, 2014.

[21] J. Galbally, et al. On-Line Signature Recognition Through the Combination of Real Dynamic Data and Synthetically Generated Static Data. Pattern Recognition, 48:2921-2934, 2015.