Pattern Recognition 45 (2012) 993–1003

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

BioSecure signature evaluation campaign (BSEC'2009): Evaluating online signature algorithms depending on the quality of signatures

N. Houmani [a,*], A. Mayoue [a], S. Garcia-Salicetti [a], B. Dorizzi [a], M.I. Khalil [b], M.N. Moustafa [b], H. Abbas [b], D. Muramatsu [c], B. Yanikoglu [d], A. Kholmatov [e], M. Martinez-Diaz [f], J. Fierrez [f], J. Ortega-Garcia [f], J. Roure Alcobé [g], J. Fabregas [g], M. Faundez-Zanuy [g], J.M. Pascual-Gaspar [h], V. Cardeñoso-Payo [h], C. Vivaracho-Pascual [h]

[a] Institut Telecom; Telecom SudParis, Department of Electronics and Physics, Evry, France
[b] Ain Shams University, Department of Computer and Systems Engineering, Cairo, Egypt
[c] Seikei University, Department of Electrical and Mechanical Engineering, Tokyo, Japan
[d] Sabanci University, Faculty of Engineering and Natural Sciences, Istanbul, Turkey
[e] TUBITAK-UEKAE, Sceintific and Technological Research Council of Turkey
[f] Universidad Autonoma de Madrid, Biometric Recognition Group-ATVS, Madrid, Spain
[g] Escola Universitaria Politecnica de Mataro, Spain
[h] Universidad de Valladolid, Department of Computer Science, Valladolid, Spain

ARTICLE INFO

Article history:
Received 8 April 2011
Received in revised form
24 June 2011
Accepted 4 August 2011
Available online 11 August 2011

Keywords:
On-line signature verification
Evaluations
Signature Categorization
Quality
Entropy

ABSTRACT

In this paper, we present the main results of the BioSecure Signature Evaluation Campaign (BSEC'2009). The objective of BSEC'2009 was to evaluate different online signature algorithms on two tasks: the first one aims at studying the influence of acquisition conditions (digitizing tablet or PDA) on systems' performance; the second one aims at studying the impact of information content in signatures on systems' performance. In BSEC'2009, the two BioSecure Data Sets DS2 and DS3 are used for tests, both containing data of the same 382 people, acquired respectively on a digitizing tablet and on a PDA. The results of the 12 systems involved in this evaluation campaign are reported and analyzed in detail in this paper. Experimental results reveal a 2.2% EER for skilled forgeries and a 0.51% EER for random forgeries on DS2; and a 4.97% EER for skilled forgeries and a 0.55% EER for random forgeries on DS3.

1. Introduction

For the last twenty years, most of the works carried out in the framework of handwritten signature verification are focused on the development and the implementation of new algorithms for online signature recognition. Such works aim at improving the performance of automatic identity verification systems based on the online handwritten signature modality.

However, even though verification systems in the literature are evaluated using publicly available databases in recent years, it is still difficult to compare the performance of such verification systems because of the differences in experimental conditions. To overcome this issue, it is important for the scientific community to conduct signature evaluation campaigns allowing an objective comparison of the algorithms with respect to each other and to standard approaches of the state-of-the-art, using the same databases and evaluation protocols.

In the past, only a few public evaluations have been organized for comparing advances in online signature verification. These include the first Signature Verification Competition (SVC) held on 2004 [18], the Signature Competition of the BioSecure Multimodal Evaluation Campaign (BMEC), held on 2007 [19], and more recently the ICDAR Signature Verification Competition, held in 2009 [20].

SVC'2004 [18] was carried out on a database of very limited size (60 people, only one session), mixing signatures of different cultural origins, captured on a digitizing tablet. The signatures in this database were not "true" signatures; indeed, the subjects were advised not to use their real signatures for privacy reasons. SVC'2004 was divided into two tasks, depending on the input features available: in Task 1, only the pen coordinates and the sample time stamps were available; in Task 2, the pen pressure and pen inclination angles (azimuth and altitude) were also available. For both tasks, the Dynamic Time Warping-based system submitted by *Sabanci University* [7] obtained the best Equal Error Rate (EER) when tested on skilled forgeries

* Corresponding author.
E-mail address: nesma.houmani@it-sudparis.eu (N. Houmani).

0031-3203/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2011.08.008

(EER=2.84% in Task 1 and EER=2.89% in Task 2). In second position, we distinguished the HMM-based systems with an EER around 6% in Task 1 and 5% in Task 2, when tested on skilled forgeries. On random forgeries, the HMM-based system submitted by *Universidad Autonoma de Madrid* [13] was the best system, with an EER of 2.12% in Task 1 and of 1.70% in Task 2.

The BMEC'2007 Signature Competition [19] was carried out in the framework of the BioSecure Network of Excellence [22,26]. It was the first signature verification evaluation on signatures captured on a mobile platform (Personal Digital Assistant, PDA) [19]. The aim of this competition was to compare the performance of different verification systems in mobile conditions, on the large BioSecure Data Set 3 (DS3) (430 people, 2 sessions) [19,22,26]. In this evaluation, we noticed that the model-based systems outperformed those based on distance approaches [19]. Indeed, the Gaussian Mixture Model-based system submitted by *EPFL* [19] obtained the best performance, when tested on both skilled and random forgeries (EER=13.43% and EER=4.03%, respectively). This winning system was followed by the HMM-based Reference System of *Telecom SudParis* [17,27,30], with an EER of 15.36% for skilled forgeries and of 4.88% for random forgeries.

The ICDAR'2009 Signature Verification Competition [20] was held in 2009, in the framework of the 10th International Conference on Document Analysis and Recognition (ICDAR'2009). This competition was carried out on the Netherlands Forensic Institute (NFI) signature database (100 people), containing both offline dataset and its corresponding online dataset acquired on a digitizing tablet. This competition was the first signature verification evaluation on offline signatures and also the first competition where offline and online signatures were combined [20]. Moreover, the competition aimed at combining expert forensic judgments with the performance of automatic verification systems by testing systems on a forensic-like dataset. For the online task, the best result was obtained by *Parascript, LLC* with an EER of 2.85%. For the offline task, the best result was obtained by *Centre for Mathematical Morphology* with an EER of 9.15%. The only system which combined both offline and online data was that of *Universidad Autonoma de Madrid*, which obtained an EER of 8.17% [20]. No information was given on the classifiers used by the submitted systems.

At the same time as ICDAR'2009 Signature Competition [20], a new evaluation campaign was organized in 2009, namely the BioSecure Signature Evaluation Campaign (BSEC'2009) [23], which was held in conjunction with the International Conference on Biometrics (ICB'2009) [21], and which is the subject of the present paper. This competition was divided into three tasks and was focused on the evaluation of online signature verification systems following new benchmarking frameworks. In the previous signature competitions, signatures were acquired with a single sensor in each competition: a digitizing tablet at SVC'2004 [18] and ICDAR'2009 [20], and a PDA at BMEC'2007 [19]. In contrast, BSEC'2009 was performed on the two existing largest databases containing the same persons, acquired with two different sorts of sensors, namely a digitizer (BioSecure Signature Corpus DS2) and a PDA touch screen (BioSecure Signature Corpus DS3) [22,26]. The DS3 corpus is indeed the first on-line signature multi-session database acquired in a mobile scenario, while the DS2 corpus was collected on a fixed platform, from the same subjects. BSEC'2009 [23] aimed at measuring the real impact of a mobile platform on algorithms' performance on these two databases. This first objective was studied in Task 1 of BSEC'2009.

The second objective of BSEC'2009 [23], studied in Task 2, was to analyze the impact of time variability on systems' performance and to assess the relative pertinence over time of the different time functions captured by the sensor [23]. It is worth noticing that there are very few works in the literature studying the impact of time variability of signatures on systems' performance. The two BioSecure databases DS2 and DS3 [22,26] are well suited to this study as they were collected in two sessions separated in time by several weeks.

Finally, a biometric system's performance is measured, in general, globally on all the available data in a database, in terms of the two types of errors that a biometric system can make, namely False Rejections and False Acceptances. This is the case of all previous signature evaluation campaigns [18,19,20]. However, it is obvious that some persons possess a signature that is easier to recognize than others. This can be related to the complexity and the stability of their signatures. Therefore, to have a better insight on the behavior of a classifier, it is wise to split the database in subsets, according to a criterion related to the difficulty of recognizing an individual. Therefore, the third objective of BSEC'2009 [23], studied in Task 3, was to evaluate the performance of different algorithms depending on the information content in the signatures, thanks to a protocol categorizing the data of both DS2 and DS3 in subsets [23]. To this end, we exploited the notion of Personal Entropy, introduced in [24,25] to categorize people depending on the quality of their signatures. Systems' performance was also measured globally on the complete databases for comparison purposes.

In this paper, we present the BioSecure Signature Evaluation Campaign BSEC'2009. As the participants did not use all combination of features in order to study the impact of time functions on systems' performance, we cannot report the results of Task 2 and we only present in this paper the results of the two major tasks, those relying on the quality of signatures. More precisely, we present the results of Task 1 studying the impact of mobile conditions, and the results of Task 3 studying the impact of information content of signatures on performance assessment. Table 1 provides a summary for BSEC'2009 and highlights the differences of this competition with respect to previous ones (SVC'2004, BMEC'2007 and ICDAR'2009) in terms of datasets used, the different tasks considered, the number of participants, and the best performance achieved.

This paper is organized as follows: Section 2 presents the two BioSecure Signature Data Sets DS2 and DS3 used for this evaluation. Section 3 describes the calculation of the Personal Entropy measure associated to a given person by means of a Writer-HMM, and how it can be used to automatically generate writer categories through a hierarchical clustering procedure. Section 4 describes the evaluation protocol and the two main tasks of BSEC'2009: Task 1 and Task 3. In Section 5, we give a brief description of the 12 submitted systems. Section 6 presents the most pertinent experimental results of Task 1 and Task 3. Finally, conclusions are stated in Section 7.

2. BioSecure signature datasets

Two datasets were used in this competition [22]. These datasets were acquired in several sites in Europe, in the framework of BioSecure Network of Excellence [22,26]: DS2 was acquired on a digitizing tablet, and DS3 was acquired on a mobile platform (PDA).[1]

For this evaluation, two development datasets of 50 people from respectively BioSecure DS2 and DS3 have been distributed to the participants. Note that for such datasets, the donors provided their own genuine signatures (not fake signatures as in SVC'2004 [18]), and the 50 people are the same in the two development datasets [23]. Besides, two other datasets containing signatures of

[1] Part of the BioSecure Signature Data Sets DS2 and DS3 are publicly available on the website of BioSecure Association [22].