https://doi.org/10.1016/j.patrec.2018.04.030

Biopen–Fusing password choice and biometric interaction at presentation level

Maria De Marsico[a,*], Federico Ponzi[a], Federico Scozzafava[a], Genoveffa Tortora[b]

[a] *Sapienza University of Rome, Via Salaria, 113, Rome 00198, Italy*
[b] *University of Salerno, Via Giovanni Paolo II, 132, Fisciano (SA) 84084, Italy*

ARTICLE INFO

Keywords:
Biometric authentication
Augmented pen
Dynamic writing recognition

ABSTRACT

The paper presents experiments with the home-made, low-cost prototype of a sensor-equipped pen for handwriting-based biometric authentication. The pen allows to capture the dynamics of user writing on normal paper, while producing a kind of password (passphrase) chosen in advance. The use of a word of any length instead of the user's signature makes the approach more robust to spoofing, since there is no repetitive pattern to steal. Moreover, if the template gets violated, this is much less harmful than signature catch. The entailed sensors are a pair of accelerometer and gyroscope and a pressure sensor. The aim is a natural yet precise interaction, that allows recognizing the user by the signals recorded while producing a specific word chosen during enrollment and possibly changed later. The pen can be exploited in a number of applications requiring user recognition, yet relieving from the need to learn complex procedures, and to undergo critical capture operations. The approach fuses the use of a kind of password, though not necessarily complex as those requested by traditional approaches, and biometric recognition. The novelty with respect to most proposals in literature is the combination of three elements at once: the matching of any handwritten text instead of user signature, the on-line capture of seven sensor signals to recognize handwriting dynamics (three from accelerometer, three from gyroscope and one from pressure sensor), and the use of normal paper instead of a digitizing tablet. Presented experiments test two different recognition techniques, implemented by two modules that can be alternatively plugged into the system. An SVM-based verification module entails to extract the most relevant features from writing dynamics, and to acquire a sufficient amount of enrolling data (30 samples per user) to train an SVM for each user. A pure Dynamic Time Warping (DTW) verification module does not require such training, and is tested using either a gallery with the same number of templates per user as those used for SVM training, or with a gallery containing a much lower number of templates per user (namely 5). Obtained results encourage further investigation of lightweight strategies for written password dynamics recognition.

1. Introduction

Authentication is the pleasure and pain of any user in the world that has to access a protected service or location. The use of passwords is dateless, as the use of special objects or signs. They were formerly used by groups, rather than by individuals. Sentinels would challenge those asking to enter an area to supply a watchword. In the Greek world, members of the Pythagorean school made a vow of secrecy, and were recognized from a 5-pointed pentagon or star tattooed on their palm. The first computer passwords were probably used at the Massachusetts Institute of Technology in the mid-1960s, when its researchers built a massive time-sharing computer called CTSS.[1] Using this computer MIT pioneered many of the milestones of computing, like e-mail, including password-based authentication too. At the beginning, a single password was sufficient to access one's virtual space and files. But computers started spreading in every-day life together with an exponentially increasing number of remote services: the number of passwords to remember grew at the same pace, together with password theft risk. This in turn calls for more and more complex passwords, difficult to crack but also to remember. Keys and cards used either as physical alternatives or add-ons often make things even more cumbersome for users. Biometric authentication, though not being invincible, has risen as a more "natural" alternative that allows users

* Corresponding author.
E-mail address: demarsico@di.uniroma1.it (M. De Marsico).

[1] Robert McMillan, "The World's First Computer Password? It Was Useless Too," http://www.wired.com/2012/01/computer-password/, January 27, 2012.

to just exploit *what they are* or *the way they behave* to attain either physical or logical access. Multibiometric systems can further enforce security by processing different physical and/or behavioral traits.

Though becoming increasingly accurate, systems designed to assure secure access to places and data are often difficult or cumbersome to use. The work in [1] explicitly underlines: "*The security research community has recently recognized that user behavior plays a part in many security failures, and it has become common to refer to users as the 'weakest link in the security chain'. We argue that simply blaming users will not lead to more effective security systems.*" Since 2000 Nielsen [2] optimistically assumes that "*in the future, security will improve through biological* [biometric] *verification mechanisms, such as fingerprint recognition or retina scanning,*"; however, the same author recognizes that " *it will take time for this infrastructure to be built (and fingerprint systems won't work for some people)*", and even more skeptically [1] concludes on the subject that "*biometric systems may be a good fit for some user-tasks-context configurations, but not all of them.*" Actually, some concerns raised in [3] in 2004 and especially related to the acquisition step, are unfortunately still valid: fingerprint readers usually require a well centered, not moving finger (and always the same!), while iris scanners can pose usability problems related to the alignment of the eye with the camera lens [4]. These problems are even more crucial with mobile biometrics, and therefore unattended acquisition, that are gaining increasing diffusion [5,6]. Further concerns regard possible privacy breach and intrusiveness [7,8], long raised and still not completely solved. A possible, only apparently obvious solution, is proposed in [9] :"*Biometric systems should have user-friendly, intuitive interfaces that guide users in presenting necessary traits.*"

This paper presents a proposal for easy-to-implement and user-friendly identity verification based on writing dynamics (defined below as on-line verification), which fuses biometrics with traditional password-based authentication, though with simplified and better user acceptable requirements. The proposed approach differs from most works in literature due to the combination of three elements at once: the matching of any (secret) handwritten text instead of user signature, the on-line capture of seven sensor signals to recognize handwriting dynamics (three from accelerometer, three from gyroscope and one from pressure sensor), and the use of normal paper instead of a digitizing tablet. The proposed system can be used also for more traditional signature-based verification, but at present it implements a recognition protocol, where user enrolling entails both choosing a password to write (possibly easier to remember than usually complex requested ones) but also capturing the associated gestural pattern, which is more difficult to imitate than the static sign. Authentication would rely on both, since the shape of sensor signals is implicitly related to the chosen word, though this may not appear in the system gallery. In addition, a further security level might be achieved by also considering the static written form of the word, adding a third barrier to intruders: knowing the password, being able to imitate its static appearance, and also imitating the way to produce it from a dynamic point of view. An added value to the approach is that theft of the enrollment templates does not entail a serious problem for the user, as it may happen for signature (it is harder than expected to have to change the way one signs) or even more for face and other physical traits (though countermeasures are represented by cancelable biometric templates).

Verification experiments are carried out following two different approaches. In the first one, a feature extraction strategy builds feature vectors (templates) from writing samples. This kind of template isused to train a Support Vector Machine (SVM) for each user. Training of a SVM exploits the features extracted from 30 samples captured during enrollment. Incoming probe samples are processed to produce a similar template which is submitted to the SVM corresponding to the claimed identity. In the second approach, a user template is made of all the unprocessed signals captured by pen sensors. Dynamic Time Warping (DTW) is used for matching the single signals and then the average similarity is returned as matching result. As for this second set of experiments, they are further divided into two groups. In one group, the templates from the same 30 samples used for SVM training are included in the system gallery (extracted from samples of enrolled users). In the second group, the set of gallery templates is dramatically reduced to 5 per user. In this way, 6 rounds of experiments are carried out, and their results averaged. Encouraging results suggest further investigations.

The paper continues as follows. Section 2 discusses the role of writing, and in particular of signature, used for authentication. Section 3 presents related work, though most prototypes are focused on signature rather than on generic handwriting recognition. Section 4 presents a multi-sensor lightweight prototype for handwriting recognition. Section 5 presents the results from the sets of experiments carried out in different conditions. Section 6 draws conclusions and sketches future work.

2. Signature for authentication

Among the early works in the literature focusing attention on biometric recognition, [10] is particularly worth mentioning. It deals with a (still valid) conceptual comparison of available means for formal identification of individuals. It classifies them as: (1) ways to merely distinguish among individuals, namely Names and Codes; (2) ways to verify individual identity, namely Knowledge-Based Identification and Token-Based Identification; and, finally, (3) biometrics, that can be used for both verification and identification. The term "biometrics" is used to refer to a variety of identification techniques which are based on some physical and *difficult-to-alienate* characteristic that entails suitable 'metric' or measurements of some kind. Clarke sketches a first taxonomy of biometric techniques (elaborating from [10]):

- appearance (e.g. the familiar passport descriptions of height, weight, skin colour, hair and eyes, visible markings; gender; race; facial hair, glasses; supported by photographs);
- social behavior (e.g. custom body-signals; voice characteristics; speech style; visible handicaps; supported by video);
- bio-dynamics (e.g. the manner in which one's signature is written; statistically-analyzed voice characteristics; keystroke dynamics, particularly in relation to login-id and password);
- natural physiography (e.g. skull measures; teeth and skeletal injuries; thumbprint, fingerprint sets and handprints; retinal scans; earlobe capillary patterns; hand geometry; DNA);
- imposed physical characteristics (e.g. collars, bracelets and anklets; brands and bar-codes; microchips and transponders).

Later on Jain simplifies the classification into two classes, i.e., physical or behavioral traits, and further elaborates on the features indicated by Clarke for human identifiers to prescribe those that a trait must present to be considered a biometrics [11]: (1) universality - every person should have the characteristic; (2) uniqueness - no two persons should be the same in terms of the characteristic; (3) permanence - the characteristic should be invariant with time; (4) collectability - the characteristic can be measured quantitatively; (5) performance - the achievable identification accuracy, the required resources to achieve an acceptable identification accuracy, and the working or environmental factors that affect the identification accuracy; (6) acceptability - to what extent people are willing to accept the system; (7) circumvention - how easy it is to fool the system.

In [12] biometric traits are further classified in *hard* (those able to support unique identification of an individual, e.g., face or fingerprints) and *soft*, defined as "*characteristics that provide some information about the individual, but lack the distinctiveness and permanence to sufficiently differentiate any two individuals.*" Actually, while traits like hair color and height may lack distinctiveness, some behavioral ones like gait may rather lack permanence, because they can be influenced by external factors, e.g. the ground slope.

Although signatures require contact with the writing instrument and an effort on the part of the user, as opposite to more acceptable contactless biometric traits like face, they have been accepted in government, legal, and commercial transactions as a method of verification. On the other hand, some authors include signature among traits that may not present sufficient stability. Even though each person has a unique style of handwriting, no two signatures of a person are exactly identical. The variations from a typical signature also depend upon the physical and emotional state of a person, therefore the identification accuracy of systems based on this highly behavioral biometric trait is reasonable but does not appear to be sufficiently high to lead to large-scale recognition. In other words, while signature can be used for verification (1:1 matching with a genuine signature of the subject) in forensics, it is hard to use it to identify a subject (1:N matching with a candidate gallery) [13]. It is possible to distinguish two groups of approaches to recognition based on signature: static and dynamic. In [14] those for which only a static visual record is available are defined as *off-line*, and those in which the pen trajectory or dynamics are captured during signature production as *on-line*. In the same work, signature verification following the former approaches, besides being long pursued [15], is considered as an art. One of the earliest realistic studies on automatic off-line signature verification dates back to the work by Nagel and Rosenfeld [16], whose experiments were carried out on scanned and digitized signature areas of real bank checks. In the same period, the first researches dealing with on-line verification also appeared in literature [17], exploiting accelerometry-related techniques. Since then, many works have investigated both strategies. It is worth underlining at this point some important conceptual differences. First, handwriting recognition in itself is to be distinguished from Optical Character Recognition (OCR). OCR is long well established in commercial applications that can recognize printed text, and always starts from a scanned image of a text that was printed mechanically. In handwriting recognition, the aim is rather to determine characters and words of text written by hand, where the writing styles of different people vary in a very significant way, so that the system must be trained to generalize across a wide variety of production patterns for the same item: even a single person does not always write in exactly the same style. Compared to this, handwriting identification, or, more frequently, signature verification for user authentication, entails to determine who wrote the text, or to verify if the writer identity is the claimed one (the claim might be implicit in the signature itself). The same difference exists between speech and speaker recognition. In general, the exploited feature extraction and classification methodologies are used differently, and training of the systems to learn relevant features relies on different guidelines. Nevertheless, the capture modalities are the same. Handwritten data is digitalized either by scanning the final product (off-line) or by special devices, e.g., a special pen on an electronic surface (on-line). On-line approaches allow capturing the precise sequence of written strokes, and therefore spatio-temporal information, whereas off-line methods only rely on the final text image. Therefore, the former ones are usually more precise for handwriting recognition. Surveys on the topics can be found in [18] and [19].

This work does not deal with signature, but with handwriting of a user-chosen "password". Many features of signatures are shared with generic handwriting. The required capturing and processing equipment is the same. Generic handwriting generally lacks the legal/forensic value of signature, but it is still useful for authentication, especially if its full dynamics are regarded, together with the knowledge of the pattern to write.

3. Related work

Resolution of data captured for handwriting recognition has dramatically increased, but the kinds of the exploited special devices have not changed much in the last years [18,20,21]. The following review of related work is especially focused on signature recognition, since most literature addresses this specific problem. A few examples of works tackle handwriting of passphrases instead. In [22], the experiments are carried out by developing generative models for a targeted user's handwriting based only on captured static (offline) samples, combined with pen-stroke dynamics. However, such statistics are learned from general population statistics, and do not account for the specific dynamics underlying the specific writing of a specific passphrase by a specific individual. An additional difference with respect to the proposal presented here is related to the capture modality: dynamics are captured by NEC VersaLite Pad and HP Compaq TC1100 pen computers, while the approach described here exploits an equipped pen and normal paper. The work in [23] introduces a multi-functional digitizing pen for the verification and identification of individuals by means of handwritten signatures, text or figures. The device records the pressure (Px, Py, Pz) and inclination (α, β) of the pen. The presented prototype is only based on pressure sensors, so that no complete dynamic pattern is taken into account. The system presented in [24] exploits a standard WACOM graphic tablet. Also the experiments in [25] on biometric recognition using handwritten text exploit a digitizing table. A more extensive review of literature presented in the following is useful to better understand the evolution of the field, and to better appreciate the features of the approach presented here that overall differentiate it from the others.

The idea of a pen computer dates back to Alan Kay, a visionary scientist who hypothesized the feasibility of personal computers since 1968, when it was about science fiction (see [26]). He guided the Learning Research group at Xerox Palo Alto Research Center in the creation in early '70 of the Dynabook prototype, an ancestor of present laptop computers. The pen computer should reproduce the pen and paper metaphor, by the digital processing of the *electronic ink*, mimicked by a position-sensing device. In early '90, this technology deserved a great attention by both all major market stakeholders, including Microsoft, IBM and Apple, and by scientific community, as testified by the large space devoted to the topic in ACM SIGCHI community [27]. The first tools used for pen-based interaction were light pens and touch screens. The former ones contain light sensors and are connected to a visual Display Unit, usually a CRT. The pen sees light from the screen and sends information to the computer via an electric pulse. The timing of the light pen and the raster scan are compared to obtain the exact (x, y) location of where the pen is pointed on the screen. The main disadvantages are some lack of precision and the fatigue of the arm due to the vertical position of the computer screen, so that these devices are suited neither for extensive handwriting nor for signature verification, that requires a better precision. Touch screens, when used with fingers, are even less precise and production requires much more space, though being more comfortable. Digitizing tablets are more precise, since it is easier to write on a horizontal surface, but the tablet and the screen are separated. Due to this, there is less interactivity, and the user must continuously change focus between

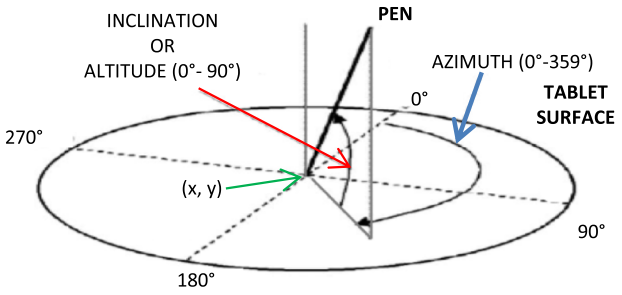

Fig. 1. Pen tip position and angles w.r.t. tablet surface.

the two. This problem is solved by notepads, that merge screen and digitizer, where a sensitive surface captures the position of a stick. At present, pen tablets are the most used devices to acquire signature for verification. The first prototypes were able to collect tip position of a tethered pen, and sometimes pressure too. The work in [14] exploits LCD writing tablets with tethered pens, using only pen tip position. More recent devices, e.g., WACOM products, collect several time series, with information recorded at fixed intervals, that include pen tip position with respect to the tablet surface, pressure applied to the pen, inclination of the pen with respect to x and y axes, or azimuth angular distance and inclination (altitude) with respect to vertical axis (see Fig. 1). Each captured signal is recorded as a time series. It is interesting to notice how the pressure plot allows to analyze the timing of pen-up/pen-down states, which are a characteristic pattern especially in characters co-articulation. In [28] on-line signature verification exploits the time series generated by pen (x, y) tip position. The dataset used in [29] contains information about tip position, pressure, azimuth and altitude angles. In [30] velocity is also computed. MCYT [31] is a dataset used for evaluation by the scientific community, which has been collected using similar hardware. It is used for experiments, e.g., in [32] and [33]. It is interesting that a low-cost prototype of a tablet used by a conventional ink pen, to be used on paper over the tablet is proposed in [34]. Actually, this methodology allows to get also the static image of the signature, and therefore to apply a multi-classifier approach. This strategy is used in the collection of recent datasets used for multi-biometric research, e.g., BiosecurID [35].

Especially for the first prototypes, the use of digitizing tablets was far from being natural and many attempts have been made to produce electronic pens being more acceptable to users and easier to integrate into existing systems. One of the earliest (1977) and most cited proposals to exploit accelerometry is the pen described in [17], following the principle that acquisition of the trajectory is an important element in signature verification. A further prototype of a pen "augmented" with sensors is presented in 1989 in [36]. Dynamic features such as changes in pen inclination and writing force are detected by various sensors attached to the pen, which is an approach similar to the one proposed in this paper. The pen inclination is estimated from its relationship with the intensity of illumination at the paper surface of a light emitted from a LED, which is measured by a reflective optical fiber sensor. The writing force in the axial direction of the pen is derived from the relationship between the strain and the applied force by using a force sensor installed at the central part of the pen. Comparison with reference data is carried out in two steps. First, simpler features are considered, such as number of pen-ups and pen-downs and maximum, minimum and average values in time of writing force and pen inclination. If comparison is promising, then a further step exploits the complete time series to get a final decision. In 2003 [37] proposes a pen equipped with two pairs of mechanical sensors that measure the horizontal and vertical movements of the ballpoint nib, and a pressure sensor on the top part of the pen, that overall produce three signals. In [38] a 3D signal approach is attempted. The electrical pen has the ability to detect the X-, Y- and Z-directional components of writing force. The Z-directional detection along the pen axis uses a conventional PZT force sensor, while a new method is proposed for the X- and Y-directional detection. This exploits a two-dimensional angle sensor to detect the 2D components of the tilt angle of the ink rod, which correspond to the X- and Y-dimensional components of writing force. In summary, electronic pens proposed in the literature are capable of "detecting" position, velocity, acceleration, pressure, pen inclination, and writing forces, using a variety of sensors: strain gauges in [39] to transduce forces and motions, magnetoelastic sensors built using wires of amorphous metal in [40], and laser diodes in [38]. Completely different proposals entail the use of a special data glove [41], or a video camera focused on the writing user, so that information is extracted from the sequence of recorded frames [42]. In the latter case, the camera focuses on a standard sheet of paper and images a common pen; the trajectory of the tip of the pen is tracked and the contact with the paper is detected. Finally, mobile devices, e.g., PDAs, may also allow signature verification [43].

4. The multi-sensor pen prototype

This paper proposes a cheap way to implement dynamic signature/handwriting-based recognition using an accelerometer, a gyroscope and a grip pressure sensor (about 20$ total for the prototype, even less on large scale production). The implemented gesture-tracking prototype system, BioPen, relies either on a Machine Learning toolkit for classification and feature extraction from time series, or on basic Dynamic Time Warping (DTW). The first goal of the pen design is to record and store as much information as possible about the hand movements performed while handling the pen for signing/writing. The second but not less important goal is to make the user feel comfortable as with a familiar everyday action. The BioPen is an ordinary medium-size ink pen, with a rubber grip equipped with special non-invasive sensors, and a bus cable which connects it to an Arduino Board. Writing is performed on common paper, as in normal signing/writing.

The proposed approach uses a gesture-tracking hardware recognition system based on 3 sensors: a 3-axis accelerometer, a 3-axis gyroscope, and a capacitive grip pressure sensor. Exploiting more signals guarantees robustness to small changes during the signature/writing procedure. Despite the user signature/passphrase is not always graphically identical, the combination of these characteristic gesture-based dynamic parameters is more flexible to variations. Moreover, they are unique and harder to reproduce by a malicious user than the static sign.

The pen is equipped with a standard MPU-6050 , an accelerometer-gyroscope-temperature sensor. This cheap and small module provides a versatile solution for motion-tracking. The module integrates a very accurate 16-bit analog to digital converter for each channel (x,y,z) for both motion sensors. The I^2C[2] interface makes handling the communication and sensor reading procedures with an Arduino board quite easy. The MPU has a built-in Digital Motion Processor, able to run complex six-axis motion algorithms (reducing the load on the main microprocessor) and a run-time calibration algorithm that provides the optimal performance to the final user. The module allows chaining with the same or additional modules (e.g., a magnetometer) in order to interact with multiple sensors at the same time. A single MPU-6050 is presently placed directly onto the pen to capture the hand position and movements.

[2] https://it.wikipedia.org/wiki/I%C2%B2C.

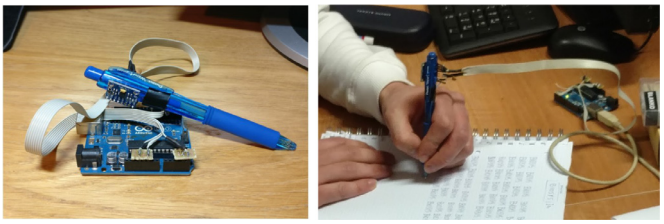

Fig. 2. The BioPen prototype and its use.

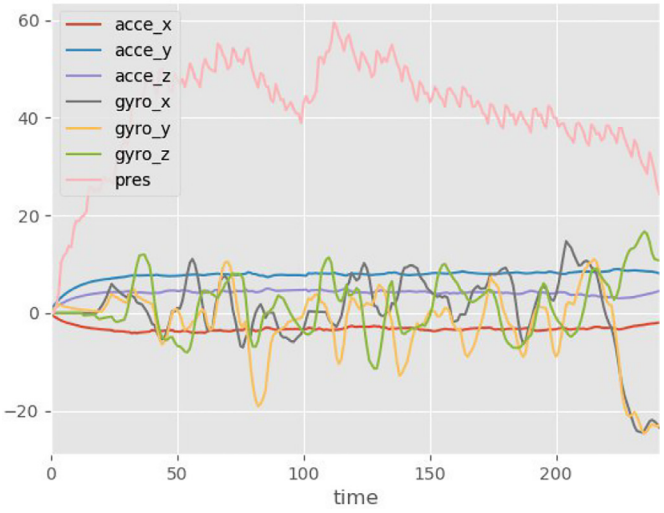

Fig. 3. The signals captured by BioPen. Time is in milliseconds (ms).

The grip pressure sensor measures the pressure applied on the pen while writing. This is a really unique and discriminant feature, rarely seen in other systems, where instead the pen to surface pressure is measured. The sensor is hand-made and consists of an aluminium foil placed under the rubber grip handle. The aluminium terminal is connected to the Arduino digital pins together with a pair of resistor-capacitor used to calibrate the operational distance threshold. In this way the sensor is able to measure the pressure on direct contact but, at its most sensitive capability, the sensor will start sensing a hand or body inches away from itself. Whenever an application calls for low to no force human touch sensing, capacitive sensing can be used. A capacitive sensor covered with paper or other insulator also acts as a fairly good (human touch) pressure sensor with an approximately logarithmic response. In this regard it may surpass force sensing resistors in some applications. A led on top of the pen just indicates the correct rotation of the pen in order to normalize the sensor signals. The hardware is connected to an Arduino board , and a suitable software handles it. Fig. 2 shows the final BioPen prototype. The application includes three main components: Acquisition, Feature extraction, and Matching module. The different modules are independent from each other, and can therefore be independently substituted with a different technology/processing choice. In particular, the feature extraction module only requires a number of time series as input, therefore it is possible to change both the hardware and the acquisition technique. In the same way, once the relevant features are extracted, the Matching module can process them according to a different strategy in order to reach a decision.

Acquisition module. A specialized Arduino code initializes and calibrates the sensors, and then transfers the acquired data at a reasonable sample rate to the processing system. A Python program parses the input from the Arduino board, and saves the output as a suitable csv file. Fig. 3 shows the time series that represent the dynamics of each acquired signal during a signing/writing session. When the prototype exploits Machine Learning, a number of user samples must be stored in the enrollment phase, in order to train the system on the individual features of each user. For the tests reported in this paper, each user was enrolled with an average of 34 samples (for a discussion about this number see Section 5.1), out of which 4 were used in turn as probes for testing and the remaining ones for training.

Feature extraction module. For Time Series data analysis, it is possible to use a recognition/matching algorithm working on the overall time series. This approach can employ, for instance, K-Nearest Neighbours (KNN) with distance measure between two series computed by Dynamic Time Warping (DTW), instead of Euclidean distance. As an alternative, feature based approaches map a time series onto another possibly lower dimensional representation. The feature extraction algorithm calculates characteristics such as the average or maximal value of the time series, the standard deviation of values, etc. The features are then passed as a feature vector to a "normal" algorithm, e.g., a Neural Network, Random Forest or Support Vector Machine if Machine Learning is exploited.

For the present prototype a feature based approach was implemented, joined with a Machine Learning process, and compared with a pair of pure DTW approaches. We adopted the FeatuRe Extraction based on Scalable Hypothesis tests (FRESH) algorithm for time series classification and regression as reported in [44]. The FRESH algorithm, as implemented in [45], automatically extracts hundrends of features from time series, like the number of peaks, the average or maximal value, summary statistics, such as maximum, variance or kurtosis, characteristics from sample distribution, such as absolute energy, whether a distribution is symmetric or the number of data points above the median, fast Fourier transformation coefficients, autocorrelation lags or mean value of the second derivative and many others. The FRESH algorithm also includes a filtering procedure for a scalable feature selection. This filtering procedure evaluates the explaining power and importance of each feature for the regression or classification tasks at hand. It is based on the well developed theory of hypothesis testing and uses a multiple test method. These tests are based on the assumption that a feature x_k is meaningful for the prediction of the binary label vector y if x_k and y are not statistically independent. As a result, the filtering process mathematically controls the percentage of irrelevant extracted features. In a second step, each feature vector is individually and independently evaluated with respect to its significance for predicting the target under investigation. The result of these tests is a vector of p-values, quantifying the significance of each feature for predicting the label/target. This vector is evaluated on the basis of the Benjamini–Yekutieli procedure [46] in order to decide which features to keep.

The feature selection process[3] returned the most relevant 169 features, making up a vector representing the user template.

Experiments also assessed an alternative strategy, where unprocessed signals from the different sensors are matched by Dynamic Time Warping (DTW). In this case, a user template is composed of all the signals acquired from the different sensors.

Matching module The Matching module works in verification mode, which is the one usually entailed by signature. It has been implemented following two alternative approaches, one exploiting Machine Learning withe templates built by feature extraction, and the other exploiting pure DTW on unprocessed signals.

[3] All procedures related to Tsfresh are available at: https://github.com/blue-yonder/tsfresh.

Support Vector Machine (SVM) is the Machine Learning approach chosen for the recognition[4]. The prototype uses a SVM with linear kernel for each user. Such SVMs must be trained before testing operations. User templates are computed using the feature extraction strategy described in the previous section. A new user has to "teach" the system how to recognize him/her by providing an adequate number of examples. Afterwards the system will be able to recognize the user again in the future and it could even possibly improve performance by reinforcement learning (Template Updating, e.g., see [47]) during each authentication session. Enrolling a new user does not require to train the overall system again, but it is sufficient to train the new corresponding SVM. Occasionally, a retraining of all the classifiers with new data can provide better performances, but this is worth being carried out only when the amount of newly acquired templates reaches a significant level.

The alternative implementation of the module exploits DTW[5]. The unprocessed signals from sensors are matched separately, and then the average similarity is returned. Both the alternatives of chaining the different signals into a single one to match, or to chain a fixed number of samples from each signal, have been discarded to be robust to possibly different lengths of the signals acquired for different samples, even if the same user is writing the same word. In particular, this would have caused an unreliable alignment of possibly sampled points.

5. Experimental results

This section presents the results from testing of the BioPen prototype with the above hardware/software settings. The presented results just refer to the recognition of the writing dynamics. Given the achieved accuracy with the same passphrase (word) for all users, security can be further improved by the choice of a personal "password" and by tuning its required complexity (though lower than that required for pure password authentication). Therefore, while a best value of EER=0.093 is achieved using SVM-based verification with the present setting, the level of security actually provided can be increased by the difficulty to guess the string to write. Even further improvement can be achieved by adding the analysis of the static written form of the text. The methods taken into account, i.e., DTW and SVM, are representative enough to point out the potential of the proposed approach.

5.1. Dataset

Information about one's way of signing is something personal than one would not easily share, or in any case is less willing to share than one's face. However, as confirmed by informal user interviews, generic handwriting is perceived as less compromising. Moreover, from an experimental point of view, asking many users to write the same word many times and then comparing the obtained templates allows to better appreciate the possible accuracy of the system taking exclusively into account pure dynamic features. Therefore, in order to evaluate the general feasibility of the approach, a single test word in block capital letters (namely the word *BIOSYS*) was used to simplify the test process. On one end, since all users wrote the same word in a common style, without any attempt to copy someone else pattern, it may seem that impostor detection testing addresses a somewhat easier problem. On the other hand, the task may be considered even harder than real signature verification, since the true signatures present much more distinctive characters. In other words, in principle, just because samples are very similar to each other, the dataset is harder to classify. In the analyzed conditions, it should be easier to cheat the system, by confusing its recognition module, therefore these conditions are particularly stressful for an automatic classification. As a consequence, in a realistic environment, a more heterogeneous dataset (different passphrases and character style together with different writing dynamics) could achieve even higher performances, being the samples and the extracted templates easier to separate. The strategy as it is can be used to implement a two-items recognition protocol. According to this protocol, user enrolling would entail both choosing a password to write but also actually writing it in a natural way, which is much more difficult to imitate. Both elements would be implicitly checked all at once during authentication (the word choice determines the dynamics to produce it). Therefore the assumption is that the presented setting is sufficient for a preliminary evaluation. Of course we limit the evaluation to the dynamics, since the evaluation of the possibility of password guess is out of the scope of this work. The number of samples is 34 on average, for each of 30 subjects. According to the central limit theorem, the number of samples used for training (30) can be deemed sufficient for a reliable inference (see [48]) if it is possible to assume clean Gaussian distributions for the different classes. Given the way the proposed system is designed and evaluated, this might not hold due to the complexities introduced by: (1) the number of training/enrolment samples, that is kept reasonably low to avoid a too much cumbersome user enrolment procedure, and (2) the use of a same passphrase for all users during system evaluation, that also allows to stress the system as in spoofing attack conditions. However, both the results achieved with SVM (Section 5.2.1) and those achieved with an even lower number of gallery samples matched by pure DTW (Section 5.2.3) are promising. This seems to testify that, notwithstanding the limitations entailed by the above introduced constraints, the selected number of samples can represent the user with sufficient accuracy. Fig. 4 shows an example of enrollment session.

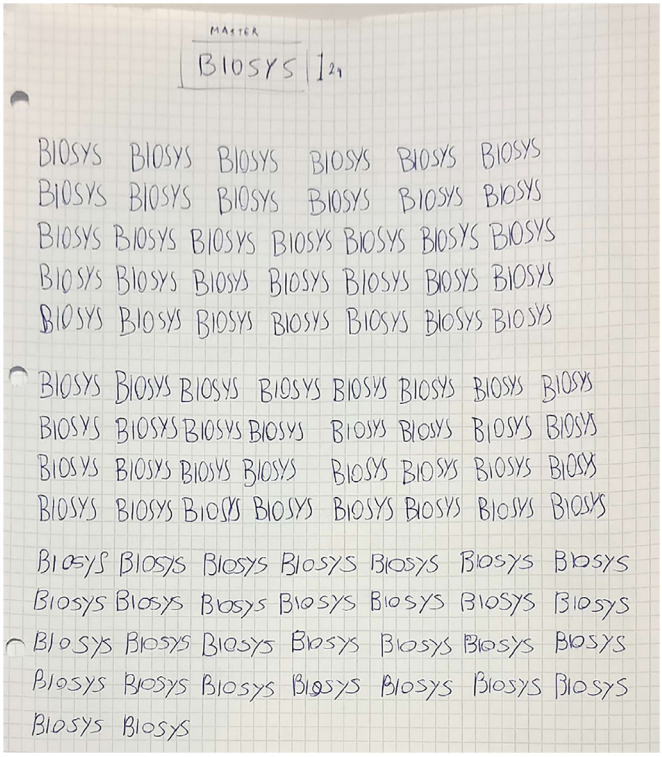

Fig. 4. An example of enrollment session.

[4] The prototype exploits the SVM implementation available from scikit-learn - Machine Learning in Python - http://scikit-learn.org/stable/modules/svm.html.

[5] The implementation of DTW is from http://alexminnaar.com/time-series-classification-and-clustering-with-python.html.

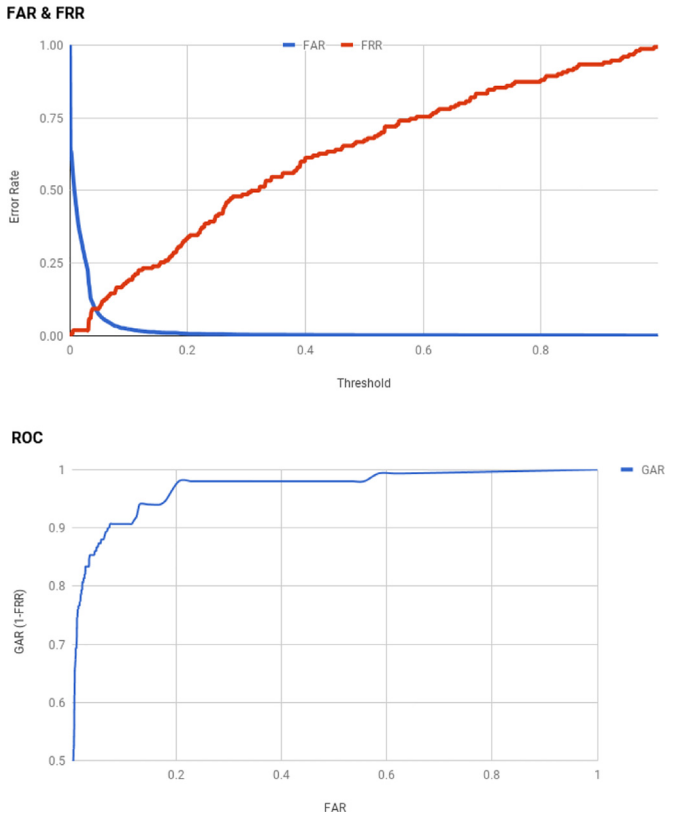

Fig. 5. FAR/FRR/EER (top) and ROC (bottom) when templates are built by feature extraction from sensor signals, and SVM is used for verification.

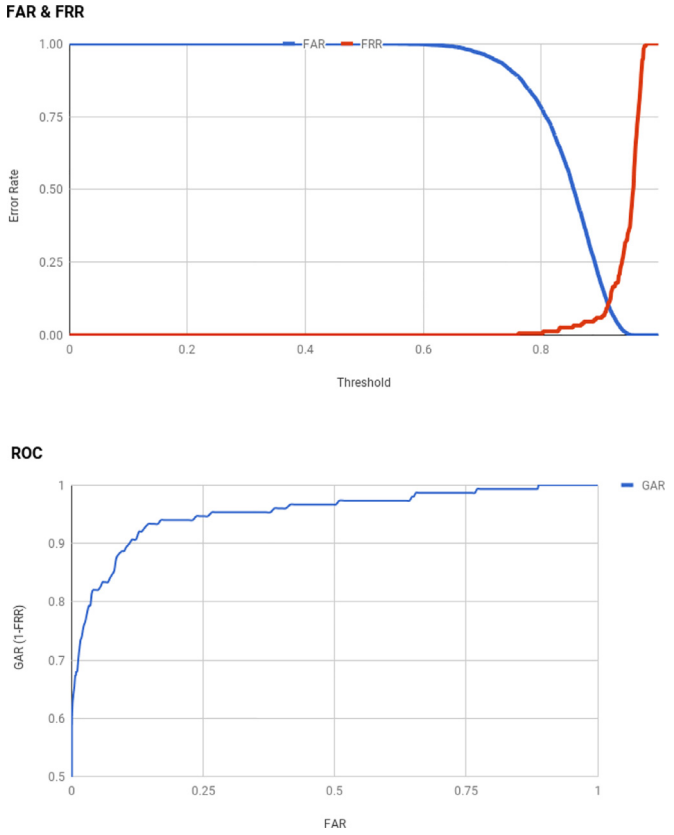

Fig. 6. FAR/FRR/EER (top) and ROC (bottom) when sets of unprocessed sensor signals are used as templates, and DTW is used for verification taking the highest similarity with gallery templates.

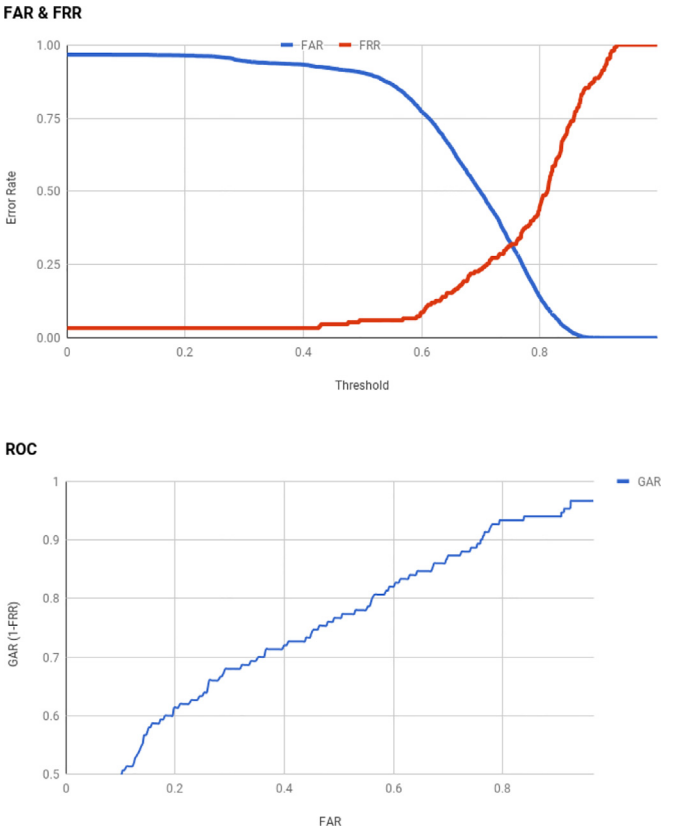

Fig. 7. FAR/FRR/EER (top) and ROC (bottom) when sets of unprocessed sensor signals are used as templates, and DTW is used for verification taking the lowest similarity with gallery templates.

5.2. Verification results

5.2.1. SVM-based verification

In order to train a bank of SVMs, each one used to recognize a single user, the captured samples were divided into training and testing sets. For each user, 4 out of the about 34 samples acquired were used for testing, and the remaining ones for training. The testing samples were used for an all-against-all matching. Each such sample was considered either as a genuine attempt or as a cheating attempt. In order to do this, it was submitted to all SVMs in parallel, to use it as a probe in a number of experiments equal to the number N of users (1 genuine and $N-1$ impostor attempts). The result returned by each SVM was compared to an acceptance threshold. For each such comparison, according to the equality/inequality of the label associated to the probe and the one of the SVM providing the result, one of four values was incremented: False Accept (FA - threshold passed but different labels), False Reject (FR - threshold not passed but same labels), Genuine Accept (GA - threshold passed and same labels), and Genuine Reject (GR - threshold not passed and different labels). The corresponding success/error rates were computed according to the number of Genuine Users (GU - once for each sample) and Impostor Users (IU - $N-1$ times for each sample), i.e., $FAR = FA/IU$, $FRR = FR/GU$, $GAR = GA/GU$, and $GRR = GR/IU$. Acceptance threshold was incremented in 0.001 steps. Fig. 5 (upper plot) shows the FAR and FRR curves, with the Equal Error Rate (EER) point (equal probability of false accept and false reject, i.e., the intersection point of FAR and FRR curves) which is usually considered as a reference to compare the performance of verification systems. For the present tests, EER=0.093, which is quite satisfying given the extremely prototypical nature of the present system, and the fact that

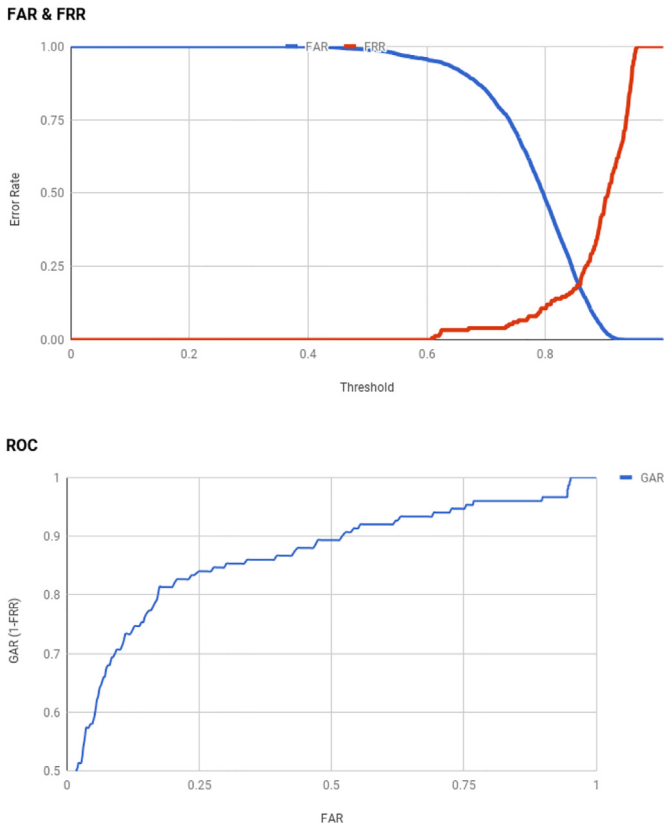

Fig. 8. FAR/FRR/EER (top) and ROC (bottom) when sets of unprocessed sensor signals are used as templates, and DTW is used for verification taking the average similarity with gallery templates.

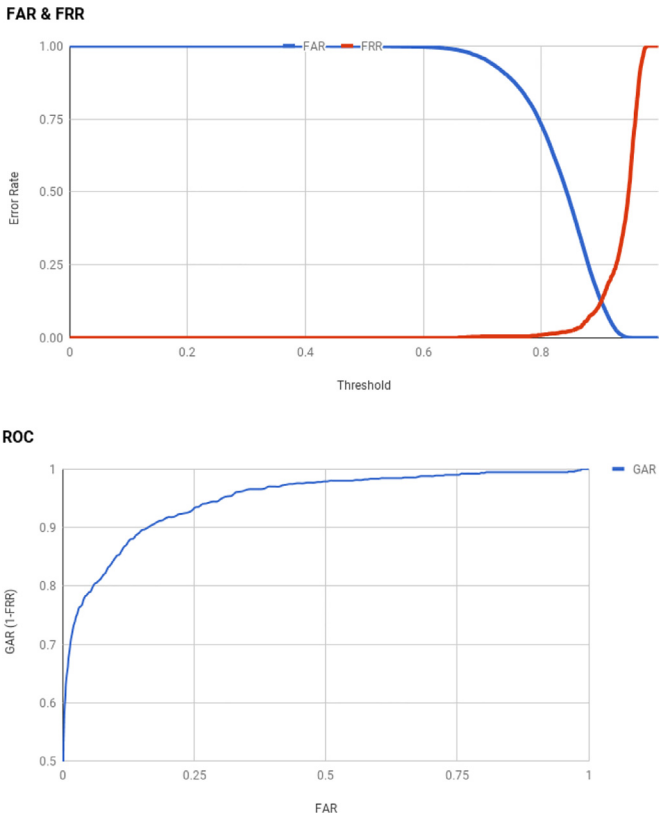

Fig. 9. Average FAR/FRR/EER (top) and ROC (bottom) computed over 6 rounds of test, when sets of unprocessed sensor signals are used as templates, and DTW is used for verification taking the maximum similarity with a gallery of 5 templates per user.

authentication also entails writing the right password. It is further interesting that FAR has a desirable trend, with the probability of false recognitions falling very quickly towards very low values. This is confirmed by the "good shape" of Receiver Operating Characteristic (ROC), quickly climbing towards the upper left corner of the plot. This produces a high value of Area Under Curve (AUC) (AUC = 0.966) as shown in the bottom part of Fig. 5.

5.2.2. DTW-based verification

In the second experiment, the composition of probe set was not modified, while the templates used for training with SVM were all included in the system gallery. In addition, there was no feature selection process, therefore a user template contained all the signals captured by the pen sensors. The aim of the experiment was to verify if, and how much, performances decrease when adopting a lighter processing strategy (no pre-processing, no-training), yet maintaining a robust enrollment phase. Matching of two user templates was carried out by applying DTW to pairs of homologous signals, and taking the average result. The acceptance threshold was set according to this new template structure and matching strategy, therefore we expected to have a different effect on similarity results. The remaining experiment parameters (performance measures and threshold step) were the same as above. In this experiment, we also tested different strategies to choose the similarity value (out of the 30 available from the user's gallery) to be compared against the threshold to determine acceptance. In practical situations, choosing the best achieved one follows the line of multiple template approaches, therefore allowing to recognize the user under different template variations and minimize false rejections. On the other hand, choosing the worst result can be used to better address possible spoofing attacks, since even if an impostor is able to counterfeit a specific sample, there will be at least one allowing to reject the attempt, therefore minimizing false accepts. Of course, the reverse of the medal is a high rate of false rejections. Finally, the average result can be used, trying to balance the above strategies. Fig. 6 shows that the behavior when considering the best similarity value out of the 30 user's samples achieves results very similar to SVM. In this case, EER = 0.104, which is not that bad given the much lighter procedure. The shape of the ROC curve and the value of AUC = 0.952 confirm this results. It is worth noticing that FAR and FRR curves have a quite common trend when using DTW (see the following figures), which is very different from SVM curves (see Fig. 5). This is due to the different similarity measure which is exploited, while EER and AUC values are correctly comparable. The results regarding choosing either the minimum similarity out of 30 gallery samples or an average value meet the expectations. The overall performance is dramatically decreased, as testified by Fig. 7, showing EER = 0.32, and confirmed by ROC (AUC = 0.709) in the figure, even if FAR decreases more quickly as the requested similarity threshold increases. On the other hand, using the average similarity value computed over the 30 user's gallery samples achieves intermediate results, as Fig. 8 shows, with EER = 0.187 and AUC = 0.866.

5.2.3. DTW-based verification with reduced gallery

The last experiment presented here was carried out by reducing the number of samples per user in the gallery, in order to reduce the cumbersome enrollment procedure. While the probe set was left again unchanged, the former gallery of 30 samples per user was divided into 6 groups of 5 samples each. Each group was used in turn as gallery, taking the best similarity as matching result, and the achieved results were averaged over the 6 rounds of test. In this experiment, both enrollment and the following pro-

Table 1
Summary of experimental results.

Verification method	EER	AUC
SVM	0.093	0.966
DTW30MAX	0.104	0.952
DTW30MIN	0.32	0.709
DTW30AVG	0.187	0.866
DTW5MAX	0.124	0.947

cessing and matching are much lighter than with SVM. Fig 9 shows the obtained results. An encouraging EER = 0.124 shows that, in principle, even dramatically reducing the number of samples per user the decrease in verification accuracy is overall acceptable. ROC curve also shows a good behavior, with a high upper left corner close to the ordinate axis. As a matter of fact, it is possible to observe a very good value of AUC (AUC = 0.947). This encourages to further investigate the use of signal processing techniques to address this kind of problem, that do not require a huge training step for each user (and therefore a huge enrolling phase). Table 1 summarizes the results obtained in the discussed experiments.

It is worth underlining once more that the recognition system was particularly stressed on the side of impostor tests, by always using the same word and the same style (capital block letters) .

6. Conclusions

The presented experiments aimed at achieving early results on the possibility to use lightweight equipment and lightweight processing to verify user identity, through the dynamics underlying the action of writing a chosen password. Verification accuracy is encouraging, though the recognition system was stressed on the side of impostor tests (the most critical ones when special security is requested) by using the same word and the same capital block letters for all samples. In addition, it is to consider that an impostor attack would further require to guess or steal the right password. In principle, just because samples are very similar to each other, the dataset is harder to classify. In other words, in the analyzed conditions, it should be easier to cheat the system, by confusing its recognition module, therefore these conditions are particularly stressful for an automatic classification. As a consequence, in a realistic environment, a more heterogeneous dataset (different passphrases and character style together with different writing dynamics) could achieve even higher performances, being the samples and the extracted templates easier to separate.

Future work will explore different approaches for recognition, either by further Machine Learning techniques or signal-based techniques, e.g., further Dynamic Time Warping variations. The advantage of the former is a lower time for the authentication, once the training has been carried out. The latter avoids a cumbersome enrollment procedure, but needs handling the variability of the acquired time series. Further experiments will be carried out by fusing the recognition of dynamics with the recognition of the written text. A dataset with real attempts of spoofing will be collected, to test the system in a more realistic (yet not necessarily more difficult) setting. As a final note on BioPen ergonomic features, all users were informally asked to express an opinion about the capture task (how easy is to use the BioPen) and the enrollment procedure. The response to the first question was positive, since the action required is extremely natural and familiar. Also, the kind of equipment does not add any special difficulty or hindering, except for the slightly higher weight than a normal pen, that could be avoided by miniaturization, and for the ribbon connecting the pen to Arduino, that could be avoided by WiFi data transfer. Of course, the huge enrollment procedure required for the Machine Learning approach was considered quite cumbersome. For this reason, we plan to focus on different approaches in the future.

References

[1] M. Sasse, S. Brostoff, D. Weirich, Transforming the weakest link a human/computer interaction approach to usable and effective security, BT Technol. J. 19 (3) (2001) 122–131.
[2] J. Nielsen, Security and human factors, Alertbox (November 2000) http://www.useit.com/alertbox/20001126.html (2000).
[3] A.S. Patrick, Usability and acceptability of biometric security systems, in: Financial Cryptography, 2004, p. 105.
[4] L. Coventry, A. De Angeli, G. Johnson, Usability and Biometric Verification at the ATM Interface, New York: ACM press, pp. 153–160.
[5] M. De Marsico, C. Galdi, M. Nappi, D. Riccio, FIRME: face and iris recognition for mobile engagement, Image Vis. Comput. 32 (12) (2014) 1161–1172.
[6] M. De Marsico, M. Nappi, D. Riccio, H. Wechsler, Mobile iris challenge evaluation (MICHE)-I, biometric iris dataset and protocols, Pattern Recognit. Lett. 57 (2015) 17–23.
[7] N.K. Ratha, J.H. Connell, R.M. Bolle, Enhancing security and privacy in biometrics-based authentication systems, IBM Syst. J. 40 (3) (2001) 614–634.
[8] V. Matyáš, Z. Říha, Biometric Authenticationsecurity and Usability, in: Advanced Communications and Multimedia Security, Springer, 2002, pp. 227–239.
[9] M.A. Sasse, Red-eye blink, bendy shuffle, and the yuck factor: a user experience of biometric airport systems, IEEE Secur. Privacy 5 (3) (2007).
[10] R. Clarke, Human identification in information systems: management challenges and public policy issues, Inf. Technol. People 7 (4) (1994) 6–37.
[11] A. Jain, L. Hong, S. Pankanti, R. Bolle, An identity-authentication system using fingerprints, Proc. IEEE 85 (9) (1997) 1365–1388.
[12] A. Jain, S. Dass, K. Nandakumar, Soft Biometric Traits for Personal Recognition Systems, in: Biometric Authentication, Springer, 2004, pp. 731–738.
[13] A. Jain, L. Hong, S. Pankanti, Biometric identification, Commun. ACM 43 (2) (2000) 90–98.
[14] V. Nalwa, Automatic on-line signature verification, Proc. IEEE 85 (2) (1997) 215–239.
[15] A. Osborn, Questioned Documents, The Lawyers' Co-Operative Publishing Co., Rochester, N. Y., 1910.
[16] R. Nagel, A. Rosenfeld, Computer detection of freehand forgeries, IEEE Trans. Comput. 26 (9) (1977) 895–905.
[17] N. Herbst, C. Liu, Automatic signature verification based on accelerometry, IBM J. Res. Dev. 21 (3) (1977) 245–253.
[18] R. Plamondon, S.N. Srihari, Online and off-line handwriting recognition: a comprehensive survey, IEEE Trans. Pattern Anal. Mach. Intell. 22 (1) (2000) 63–84.
[19] W. Hou, X. Ye, K. Wang, A survey of off-line signature verification, in: Intelligent Mechatronics and Automation, 2004. Proceedings. 2004 International Conference on, 2004, pp. 536–541.
[20] R. Mohammed, R. Nabi, M. Sardasht, R. Mahmood, R. Nabi, State-of-the-art in handwritten signature verification system, in: Computational Science and Computational Intelligence (CSCI), 2015 International Conference on, 2015, pp. 519–525.
[21] D. Impedovo, G. Pirlo, Automatic signature verification: the state of the art, IEEE Trans. Syst. Man Cybernet. Part C (Appl. Rev.) 38 (5) (2008) 609–635.
[22] L. Ballard, D. Lopresti, F. Monrose, Evaluating the security of handwriting biometrics, Tenth International Workshop on Frontiers in Handwriting Recognition, Suvisoft, 2006.
[23] C. Hook, J. Kempf, G. Scharfenberg, A novel digitizing pen for the analysis of pen pressure and inclination in handwriting biometrics, Biom. Authentic. (2004) 283–294.
[24] M. Bashir, F. Kempf, Advanced biometric pen system for recording and analyzing handwriting, J. Signal Process. Syst. 68 (1) (2012) 75–81.
[25] E. Sesa-Nogueras, M. Faundez-Zanuy, Biometric recognition using online uppercase handwritten text, Pattern Recognit. 45 (1) (2012) 128–144.
[26] C. Davidson, The man who made computers personal, New Sci. 138 (1878) (1993) 32–35.
[27] A. Meyer, Pen computing: a technology overview and a vision, ACM SIgCHI Bull. 27 (3) (1995) 46–90.
[28] T. Ohishi, Y. Komiya, T. Matsumoto, Online signature verification using pen-position, pen-pressure and pen-inclination trajectories, in: Pattern Recognition, 2000. Proceedings. 15th International Conference on, 4, 2000, pp. 547–550.
[29] M. Faundez-Zanuy, On-line signature recognition based on vq-dtw, Pattern Recognit. 40 (3) (2007) 981–992.
[30] M. Soltane, Product of likelihood ratio scores fusion of dynamic face, text independent speech and on-line signature based biometrics verification application systems, Med. J. Model. Simul. 4 (15) (2015) 36.
[31] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J.-J. Igarza, C. Vivaracho, et al., Mcyt baseline corpus: a bimodal biometric database, IEE Proc. Vis. Image Signal Process. 150 (6) (2003) 395–401.
[32] C. Vivaracho-Pascual, M. Faundez-Zanuy, J. Pascual, An efficient low cost approach for on-line signature recognition based on length normalization and fractional distances, Pattern Recognit. 42 (1) (2009) 183–193.
[33] J. Pascual-Gaspar, M. Faundez-Zanuy, C. Vivaracho, Fast on-line signature recognition based on vq with time modeling, Eng. Appl. Artif. Intell. 24 (2) (2011) 368–377.

[34] D. Hamilton, J. Whelan, A. McLaren, I. MacIntyre, A. Tizzard, Low cost dynamic signature verification system(1995).
[35] J. Fierrez, J. Galbally, J. Ortega-Garcia, M.R. Freire, F. Alonso-Fernandez, D. Ramos, D.T. Toledano, J. Gonzalez-Rodriguez, J. Siguenza, J. Garrido-Salas, et al., Biosecurid: a multimodal biometric database, Pattern Anal. Appl. 13 (2) (2010) 235–246.
[36] H. Taguchi, K. Kiriyama, E. Tanaka, K. Fujii, On-line recognition of handwritten signatures by feature extraction of pen movements, Syst. Comput. Jpn. 20 (10) (1989) 1–14.
[37] O. Rohlík, P. Mautner, V. Matousek, J. Kempf, Hmm based handwritten text recognition using biometrical data acquisition pen, in: Computational Intelligence in Robotics and Automation, 2003. Proceedings. 2003 IEEE International Symposium on, vol. 2, IEEE, 2003, pp. 950–953.
[38] H. Shimizu, S. Kiyono, T. Motoki, W. Gao, An electrical pen for signature verification using a two-dimensional optical angle sensor, Sens. Actuators, A 111 (2) (2004) 216–221.
[39] H. Crane, J. Ostrem, Automatic signature verification using a three--axis force-sensitive pen, IEEE Trans. Syst. Man Cybern. (3) (1983) 329–337.
[40] A. Zhukov, M. Vázquez, J. García-Beneytez, Magnetoelastic sensor for signature identification based on mechanomagnetic effect in amorphous wires, Le J. de Phys. IV 8 (PR2) (1998) Pr2–763.
[41] S. Sayeed, R. Besar, N.S. Kamel, Dynamic signature verification using sensor based data glove, in: Signal Processing, 2006 8th International Conference on, vol. 3, IEEE, 2006.
[42] M. Munich, P. Perona, Visual input for pen-based computers, IEEE Trans. Pattern Anal. Mach. Intell. 24 (3) (2002) 313–328.
[43] M. Martinez-Diaz, J. Fierrez, R. Krish, J. Galbally, Mobile signature verification: feature robustness and performance comparison, IET Biom. 3 (4) (2014) 267–277.
[44] M. Christ, A.W. Kempa-Liehr, M. Feindt, Distributed and parallel time series feature extraction for industrial big data applications, arXiv:1610.07717 (2016).
[45] Tsfresh toolkit, 2017.
[46] Y. Benjamini, D. Yekutieli, The control of the false discovery rate in multiple testing under dependency, Ann. Stat. (2001) 1165–1188.
[47] U. Uludag, A. Ross, A. Jain, Biometric template selection and update: a case study in fingerprints, Pattern Recognit. 37 (7) (2004) 1533–1542.
[48] R.A. Johnson, G.K. Bhattacharyya, Statistics: Principles and Methods, 1996.