Score optimization and template updating in a biometric technique for authentication in mobiles based on gestures

J. Guerra-Casanova*, C. Sánchez- Ávila, A. de Santos Sierra, G. Bailador del Pozo

Centro de Domótica Integral (CeDInt-UPM), Universidad Politécnica de Madrid, Campus de Montegancedo, 28223 Pozuelo de Alarcón, Madrid, Spain

A B S T R A C T

This article focuses on the evaluation of a biometric technique based on the performance of an identifying gesture by holding a telephone with an embedded accelerometer in his/her hand. The acceleration signals obtained when users perform gestures are analyzed following a mathematical method based on global sequence alignment. In this article, eight different scores are proposed and evaluated in order to quantify the differences between gestures, obtaining an optimal EER result of 3.42% when analyzing a random set of 40 users of a database made up of 80 users with real attempts of falsification. Moreover, a temporal study of the technique is presented leeding to the need to update the template to adapt the manner in which users modify how they perform their identifying gesture over time. Six updating schemes have been assessed within a database of 22 users repeating their identifying gesture in 20 sessions over 4 months, concluding that the more often the template is updated the better and more stable performance the technique presents.

Keywords:
Mobile authentication
Template updating
Gestures
Biometrics
Security
Accelerometers

1. Introduction

Improving the level of security in mobile phones is a crucial task in order to keep the personal information included on these devices safe, such as the agenda, or the mail, and also to be able to carry out sensitive operations on the Internet where the correct authentication of the user trying to access is critical.

At present there are also different approaches bringing together biometric techniques, physical or behavioral (Jain et al., 2007), and mobile phones to enhance their security (Chirillo and Blaul, 2003). In Tao and Veldhuis (2006) and Ijiri et al. (2006) people are authenticated by the recognition of their face through the camera of a mobile phone. In ho Cho et al. (2006) and Jeong et al. (2005) authentication is provided by means of iris scanning and in Shabeer and Suganthi (2007) the characteristic biometric features are the voice and the fingerprint. Moreover, in Clarke and Furnell (2007) users are recognized by keystroke analysis.

In this article, we work on a novel biometric technique adapted to mobile phones with an embedded accelerometer, where users are authenticated when they perform an identifying gesture correctly (in-air signature), created by them, by holding a mobile telephone in their hand. This biometric technique stands out because of the high acceptability of the users, since it is quite similar to a handwritten signature and the suitable resistance to fraud when imposters try to repeat an authentic in-air signature.

Indeed, in Guerra-Casanova et al. (2010) an Active impostor test was presented, obtaining an Equal Error Rate of 2.5% when 40 people repeated their own in-air signature and three fraudsters tried to forge each of them by studying the performance of the authentic gestures filmed on video. This result was obtained when a sequence alignment method was carried out to analyze the signals involved in the process. In this article, we use again sequence alignment but we focus on the study of different scores to quantify the differences between the two gestures, trying to reduce the Equal Error Rate of the system. Actually, four different scores and a strategy of normalization are proposed, deriving in eight dissimilar experiments to be evaluated.

As well as finding an optimal score to compute the differences between gestures, it is important to assess the behavior of the technique over time. All behavioral biometric techniques, like the one proposed in this article, involve the fact that users may modify their behavior over time. As a consequence, a continuous biometric authentication may be required (Solami et al., 2010) or a template updating procedure (Li et al., 2008; Rattani et al., 2009).

In the context of the biometric technique based on identifying gestures, a solution based on continuous authentication is not appropriate since users are not making their gesture at every moment. However, template updating may offer a suitable solution to adapt templates and variances to the making of of gestures. Actually, this is a common practice in other biometric techniques where the biometric characteristic may change over time, as in

hand recognition (Amayeh et al., 2009), fingerprint (Freni et al., 2008) or face (Singh et al., 2009; Marcialis et al., 2008; Rattani et al., 2008).

Summarizing, this article focuses on two main objectives:

- Finding an optimal configuration of the algorithm in order to reduce as much as possible the errors when the system denies access to authentic users as well as granting access to an impostor.
- By studying the strength of the technique over time, proposing and evaluating different template updating strategies in order to adapt the template of users to how they modify unconsciously the way in which they perform their identifying gesture.

According to this, the article is divided into the following sections: In Section 2 the mathematical method proposed to compare two gestures is introduced. In this description, the eight different scores studied to quantify the differences between gestures are explained. Next, in Section 3 the results of applying each score to analyze the error rates in a database of 40 users with real falsifications are presented. As a consequence of the results of this section, the score with the lowest error rate is selected as the optimal and used in the subsequent sections. After that, Section 4 includes a temporal study when no updating is carried out as a means of confirming the necessity of updating. Then, six different updating strategies, with three variations each, are presented and evaluated. Finally, in Section 5 the conclusions of the work are provided.

2. Mathematical analysis method

When a user performs his/her identifying gesture in the air by holding a telephone with an embedded accelerometer in his/her hand, the feature extraction system obtains the accelerations of the gesture in the three axes sampled at frequency rate of 50 Hz, enough to distinguish between different repetitions with a compromise of quality and consumption time (Kela et al., 2006; Mäntyjärvi et al., 2004).

Thereby, the analysis of the biometric system should deal with acceleration signals, considering the following characteristics when the same identifying gesture is repeated:

- The beginning of the gestures does not match.
- There are peaks of acceleration more pronounced than others.
- Gestures do not take the same amount of time to be performed.
- In certain parts of the performance of the gestures, transitions are faster or slower.

According to these particularities of the acceleration signals, an analysis method based on sequence alignment is proposed to solve the differences between two repetitions of the same gesture performed by the same user but keeping significant the contrast when the gesture is made by other user (Durbin et al., 2006).

In this approach there is no feature normalization process, since how users hold the mobile and make the movements to perform their identifying gesture is very valuable information. Indeed, the goal of this technique is not to recognize the identifying gesture but the person who makes it, therefore, information such as how the user naturally holds the device and move it through the air is very important and would be discarded if the algorithm were independent to orientation. According to this, the biometric technique proposed is oriented dependant but the slight differences in holding the device or making the gesture are corrected by the algorithm described below.

The general method of analyzing two acceleration signals is explained in this Section. As specified by the scope of this article, the proposed scores to measure the differences between performances of signatures are defined throughout the description of the mathematical method. It is remarkable that the algorithm explained considers only the acceleration signals of one axis, so when two gestures are compared, three executions of the algorithm should be run, and obviously, three values of each score are obtained. The final value of the score measuring the differences at the gesture level is calculated by the average of the scores obtained at the axis level.

Therefore, the algorithm proposed tries to find the best global alignment between two signals of acceleration v, w. For this reason, a matrix S of punctuations is created and filled dynamically following Eq. (1):

$$s_{i,j} = \max \begin{cases} s_{i,j-1} + h \\ s_{i-1,j-1} + \Delta \\ s_{i-i,j} + h \end{cases} \quad (1)$$

In this equation, it is observed that:

- The overall punctuation is increased by penalty h when the punctuation on a point (i, j) of the matrix comes from its vertical or horizontal neighbour. These kinds of movements on matrix S will correspond to the introduction of a zero value in that point of sequence v (if vertical movement) or w (if horizontal) in order to find the optimal alignment of both sequences. Furthermore, h should comply with Eq. (2) so that the algorithm works properly.

$$h < 0.5 \quad (2)$$

- The punctuation of a diagonal movement depends on the value of a fuzzy function Δ, representing to what extent two points v_i, w_j are similar. Δ follows Eq. (3), where σ is a parameter used to normalize the difference between two points into a gaussian (de Santos Sierra et al., 2008)

$$\Delta = e^{-\frac{(w_j - v_i)^2}{2\sigma^2}} \quad (3)$$

Consequently, matrix S is completed depending on the punctuation filled in previous point of the signals and whether the two points of the sequences compared are more similar than the penalty of including a gap to find the best global alignment.

At this point, the first two scores are proposed in order to quantify the differences between signals:

- The first score ψ_1 is defined by Eq. (4) representing the value of the last point of the matrix of punctuations S, which is equivalent to the maximization of the score proposed in 1 over all the points of the sequences.

$$\psi_1 = S(m, n) \quad (4)$$

- The second score ψ_2 is defined as the number of gaps introduced into the best alignment of the sequences, as follows in Eq. (5)

$$\psi_2 = \#Gaps \quad (5)$$

At this point, the general analysis method carries out a backtracking algorithm in order to find the optimally aligned signals v' and w'. This algorithm consists of discovering the path to travel on matrix S from $S(m, n)$ to $S(1, 1)$ depending on the expression selected in Eq. (1) to calculate each point of the path. Any vertical or horizontal movement means including a zero in that point of v or w. Consequently, v' and w' are obtained by including some zeros in particular points in order to be aligned optimally.

These zero values are interpolated and thereafter a distance is calculated in order to measure the differences between the already

aligned signals. Two different distances have been proposed, leading to scores ψ_3 and ψ_4:

- The third score ψ_3 is obtained as the Euclidean distance between the aligned signals as described in Eq. (6):

$$\psi_3 = \sum_i (v'_i - w'_i)^2 \tag{6}$$

- The forth score ψ_4 represents the sum of the differences between the aligned signals in absolute value, Eq. (7)

$$\psi_4 = \sum_i |v'_i - w'_i| \tag{7}$$

In all the scores presented previously, the length of the signatures involved is not considered. However, it might seem quite obvious that the longer a signature is the more differences may appear. This behavior is studied by the normalization in the length of all the scores. Therefore, the normalization factor L is obtained as the average of the length of the signals in comparison, and used to normalize the score as described in the following Equations:

$$\psi_5 = \frac{\psi_1}{L} \tag{8}$$

$$\psi_6 = \frac{\psi_2}{L} \tag{9}$$

$$\psi_7 = \frac{\psi_3}{L} \tag{10}$$

$$\psi_8 = \frac{\psi_4}{L} \tag{11}$$

As gesture samples consist of three signals of acceleration (one for each axis), when two gestures are compared, three implementations of the algorithm are required and three punctuations of ψ_i are obtained, one for each axis. Ψ_i denotes the average of the three punctuations obtained when analyzing all the signals of each axis for each score ψ_i and represents the quantification of the differences between the two gestures inspected.

A user who enrolls in the system should repeat his/her identifying gesture three times. Afterwards, each pair of gestures is analyzed, obtaining three resulting values of Ψ_i. The average of the comparison of each pair of the three performances of gestures at enrollment is symbolized as μ_i in accordance with Eq. (12). This value is stored with these signals as the identifying gesture template of the user.

$$\mu_i = \frac{\Psi_i^{12} + \Psi_i^{13} + \Psi_i^{23}}{3} \tag{12}$$

When an already enrolled user wishes to access the system, he/she should carry out his/her identifying gesture once. Then, this sample is compared with the three gestures performed at the enrolling phase, obtaining three values Ψ_i^j (j means the sample of the template with which it has been compared with). The final value Ψ_i is calculated as the average of each Ψ_i^j and represents to what extent the gesture made is similar to all the samples in the template. The lower it is, the more similar the performance of the gesture is in relation to the template.

If Eq. (13) is complied with, the user would access the system. Otherwise, he/she would be rejected. Obviously, the higher the threshold θ_i is, the more falsification attempts would be accepted into the system but the less original users would be rejected, and vice versa.

$$\frac{\Psi_i}{\mu_i} < \theta_i \tag{13}$$

3. Score selection

The evaluation of the results of the experiments in this Section is carried out through the analysis of a database of 80 users who have repeated their identifying gesture seven times in the air holding an iPhone in their hand used to perform the gesture. All of these gestures have been filmed on video. From the study of these records, each original gesture has been tried to be imitated by three different people (8 trials each), representing impostors attempting to forge the biometric system. Consequently, the database of gestures considered in this article is made up of 560 samples of original in-air signatures and 1920 real attempts at falsification.

In this article we present an evaluation of the eight scores proposed in the previous section in order to find the optimal and analyze the fraud resistance of this biometric technique based on gestures when an impostor tries to falsify an authentic in-air signature of someone else.

The assessment of the fraud resistance is performed in two steps:

1. Firstly, the Equal Error Rate is obtained from a subset of the samples of the database (40 users, which means half of the in-air signatures of the database). The EER has been calculated 10 times per experiment. Each of them has been carried out with a different and random division of the database in order to make the results independent to the gestures in the database. For each score evaluated, 200 different configurations of the parameters σ and h of the algorithm have been tested (h from 0 to 0.5 with intervals of 0.025 and σ from 0 to 1 with intervals of 0.1). The configuration for each score achieving the lowest EER value has been selected. Each EER was obtained as follows for each score ψ_i (Wayman et al., 2004):
   - *Template creation:* Three samples of each in-air signature are considered as the template of the user. Then, μ_i is calculated as explained in Section 2.
   - *Analysis of original samples:* The remaining four original samples of each gesture are used to evaluate whether the system grants access to the trutful users or, on the contrary, denies access to original users. For each original trial, Ψ_i/μ_i is obtained when comparing the accessing gesture with the three gestures of the original user template.
   - *Analysis of falsified samples:* All the falsifying attempts trying to access the system will be used to evaluate whether the system is able or not to reject impostors. For each falsification trial, Ψ_i/μ_i is also obtained.
   - *Obtention of False Acceptance Rates* (FAR(θ_i)) *and False Rejection Rates* (FRR(θ_i)): FAR(θ_i) and FRR(θ_i) are obtained in terms of θ_i as the % of original samples that are over θ_i in case of False Rejection Rates and the % of falsified samples that are under θ_i in case of False Acceptance Rates. It is proven that when θ_i is very low, most falsifications are rejected but also some authentic access are not accepted. On the other hand, the higher the θ_i, the more original access are allowed but also the more falsifications are granted. FAR(θ_i) and FRR(θ_i) are obtained for values of θ_i from 0 to max(Ψ_i/μ_i) in 10,000 steps.
   - *Obtention of Equal Error Rate (EER) and the optimal threshold* (θ_{EER}): EER is defined as the value of the error when the False Acceptance Rates are equal to the False Rejection Rates. The value of the threshold at the intersection of both rates θ_{EER} is the optimal threshold value the system should implement in order to get as few errors as possible.
2. Secondly, a testing phase is carried out in order to evaluate to what extent the error rate results of the optimal configuration depend on the set of gestures of the database selected. Therefore, for each of the 10 repetitions of the algorithm, the samples of users that were not used in obtaining the EER, are used to cal-

Table 1
Results for each score.

Score	Optimal EER	Threshold θ_{EER}	FAR(θ_{EER})	FRR(θ_{EER})
ψ_1	5.46 ± 1.61	0.96 ± 0.01	8.26 ± 6.49	10.19 ± 6.65
ψ_2	7.87 ± 1.01	1.26 ± 0.02	9.05 ± 3.00	9.30 ± 2.76
ψ_3	3.95 ± 0.96	1.36 ± 0.03	4.01 ± 2.41	5.03 ± 2.43
ψ_4	3.73 ± 0.86	1.42 ± 0.04	6.38 ± 2.71	5.66 ± 1.70
ψ_5	4.60 ± 1.76	0.96 ± 0.02	6.05 ± 3.45	6.23 ± 3.32
ψ_6	7.77 ± 1.49	1.37 ± 0.03	11.81 ± 3.89	8.04 ± 1.85
ψ_7	4.67 ± 1.51	1.43 ± 0.06	6.76 ± 3.31	4.45 ± 2.78
ψ_8	3.42 ± 1.22	1.54 ± 0.07	3.67 ± 3.11	4.10 ± 2.63

culate False Acceptance Rate and False Rejection Rate in respect to the threshold θ_{EER} previously obtained:

- *False Acceptance Rate:* FAR is obtained as the % of the falsifying samples that are accepted in the system when the value of the threshold is set to θ_{EER}. This is equivalent to the percentage of fraudulent samples complying with $\Psi_i/\mu_i < \theta_{EER}$.
- *False Rejection Rate:* FRR is calculated as the % of the authentic samples that are rejected in the system when the value of the threshold is set to θ_{EER}, which is analogous to the percentage of original samples fulfilling $\Psi_i/\mu_i > \theta_{EER}$.

Summarizing, for of each configuration and each score, the following process has been carried out: firstly, by dividing the database into two halfs, secondly, calculating the EER and the optimal threshold for the first half of the in-air signatures and finally, evaluating FAR and FRR with the second half of the samples of the database and the previously obtained optimal threshold. It is highlighted that each division of the database has been carried out randomly, and experiments for each configuration and each score have been repeated 10 times in order to make the results independent of the subset selected.

Following all these considerations, the eight scores proposed in Section 2 have been evaluated, obtaining the results presented in Table 1. In this table, the results of the configuration for each score with the lowest average EER are presented. Specifically, each row of the table represents the optimal result for each score, and presents the following values:

- *Optimal EER:* The lowest value of EER obtained (Average and standard deviation of the 10 repetitions of the algorithm with different and random division of the database)
- *The value of* θ_{EER}*:* the threshold value corresponding to the intersection point EER has been obtained from (Average and standard deviation).
- *False Acceptance Rate:* Percentage of accepted forgeries of the samples not used to calculate EER, setting the threshold to θ_{EER} (Average and standard deviation).
- *False Rejection Rate:* Percentage of rejected truthful access of the samples not used to calculate ERR, setting the threshold to θ_{EER} value (Average and standard deviation).

From Table 1 it may be concluded that:

- Normalization may improve or worsen the results of the algorithm. In particular, the best result is achieved when normalizing.
- The lowest Equal Error Rate has been obtained when quantifying the differences between two performance of gestures with the score ψ_8. This result has been obtained within a configuration of the parameters of the algorithm of $h = 0.2$ and $\sigma = 0.65$.
- When evaluating with other gestures of the database EER has been obtained from, FAR and FRR are higher than EER.
- FAR and FRR of scores are relatively close to the EER value, which means that the closer they are the more optimally the system behaves when gestures not used to calculate the optimal θ_{EER} are taken into consideration. In particular, FAR and FRR of score ψ_8 are quite close to the respective EER value.

Finally, the experiments described in the following Section are carried out within the optimal configuration found in this section, which means using score ψ_8, $h = 0.2$, $\sigma = 0.65$ and the value of the threshold $\theta_{EER} = 1.54$.

4. Updating scenarios: proposal and evaluation

From the previous Section, it is concluded that a high performance in terms of Equal Error Rate could be obtained when authenticating users with gestures in the air while holding a mobile device in comparison with real falsification attempts. However, it should be taken into consideration that users are not able to repeat their gestures exactly over time, since they constantly modify how they perform them. Consequently, False Rejection Rate would increase due to the performance of the gestures of the users through the time, and obviously, Error Rates of the system would increase as well. In this context, a template updating method is needed in order to adapt the template to the variance of the performance of the gestures.

According to this, an evaluation of the behavior of the technique when users access the system with their identifying gesture in many different sessions over a long period of time is presented in this section. For this purpose, a "permanence" database has been developed consiting of 22 users repeating their identifying gesture, created by them, five times in 20 sessions spread over four months. Each session of each user has been obtained in intervals from 1 to 5 days. Moreover, sessions 10 and 11 of each user have been separated by at least one month.

In this section, firstly, an evaluation of the samples of the database when no updating strategy is applied is presented (Scenario 0), in order to confirm the need to update when users access the system over time. Furthermore, this section introduces the results of the evaluation of the samples in the "permanence" database in six groups of template updating scenarios, with different updating intervals. For each group of scenarios, the study of the updating strateg yis included in accordance with the following cases:

- (3 M) The three samples that make up the template are modified alternatively: On each updating process, the oldest of the three samples of the template is modified by the new one.
- (2 M) Two of the samples of the template are changed in turns: On each updating process, the oldest of two of the samples of the template is modified by the new one whereas one of the template samples is never substituted.
- (1 M) Only one of the three template samples is replaced: On each updating process, only one of the samples of the template, and always the same, is modified whereas the other two, obtained at enrollment, are never varied.

Taking this into consideration, the groups of template updating scenarios assessed in this article are:

- *Scenario 0:* No updating.
- *Scenario 1:* "Always updating": Updating the template on each successful access of the user to the system.
- *Scenario 2:* "One-out-of-two updating": Updating the template on each one out of two successful access of the user.
- *Scenario 3:* "One-out-of-three updating": Updating the template on each one out of three successful access of the user.
- *Scenario 4:* "In-1-session updating": Updating the template once on each session.

- *Scenario 5:* "In-2-session updating": Updating the template once on each two sessions.
- *Scenario 6:* "In-3-session updating": Updating the template once on each three sessions.

In all the experiments introduced in this section, the three first attempts of each user at the first sessions are considered as the gestures performed at enrollment phase, building the initial template of each user which is modified following the instructions of each updating strategy.

It is remarkable that when an updating process is carried out, not only the substitution of one of the samples of the template by the successful access is performed, but also, μ_8 is recalculated. Therefore, as the template has been completely modified, False Acceptance Rate (FAR)) should be determined again. According to this, in all the experiments, FAR is calculated each time each template is updated, storing the results separately for each session. In addition to this, False Rejection Rate (FRR) is also obtained separately by sessions, from the results of the access to the system of the original users in comparison with their templates in the state of updating of that precise session. Note that False Acceptance Rate and False Rejection Rate are obtained using the value of the optimal threshold obtained in the previous Section:

- False Acceptance Rate is calculated as the % of impostor samples complying with $\Psi_8/\mu_8 < 1.54$.
- False Acceptance Rate is the % of original samples that $\Psi_8/\mu_8 > 1.54$.

In this manner, the results of these experiments are independent of the optimization of the system since they have not been used to obtain the Θ_8 value.

According to this, each scenario presents the following results in order to evaluate the updating strategy and the improvement it introduces with respect to the system in Scenario 0 when no template updating is carried out:

- False Acceptance Rate and False Rejection Rate per session: For each scenario, FAR and FRR are obtained by using the samples of each session separately. Thus, for each session, the original samples of the session are used to obtained FRR and update the template. On each updating of the template, FAR is calculated from the comparison of the impostor samples with the updated templates. (As there are no falsifying attempts in this database, the impostor samples used for each in air signature are the original samples of the rest of the users). These results are represented in different figures (one for each scenario), symbolizing the behavior of the error rates over time for each updating proposal.
- Mean and standard deviation of FAR and FRR: In order to deduce which scenario provides better results, an overall analysis should be carried out. Accordingly, the mean and the standard deviation of FAR and FRR are calculated as a means of comparing the average and the variance of the errors in a general manner.

4.1. Scenario 0: No updating

In this scenario, the results when no template updating strategy applied are presented in order to evaluate, firstly, the convenience and necessity of an updating method.

Therefore, False Rejection Rate is calculated by considering all the original access of the users at different sessions (except those used for enrollment). Furthermore, False Acceptance Rate is also obtained by using the samples of the rest of the users as impostor attempts. Both rates are evaluated with a value of $\Theta_8 = 1.54$, as obtained in the previous Section.

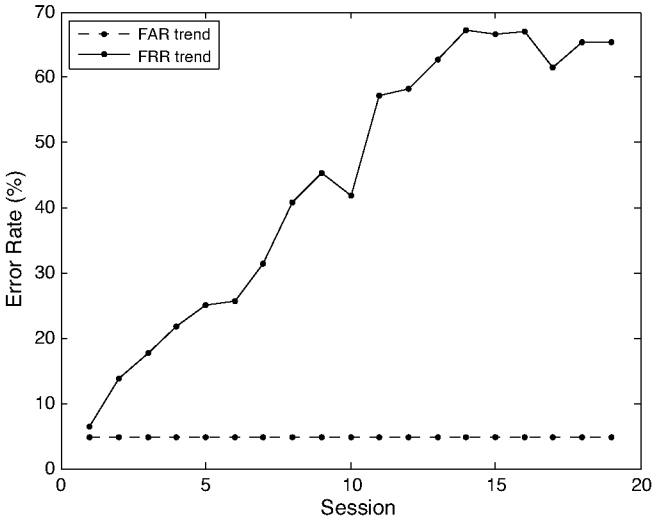

Fig. 1. Scenario 0: False Acceptance and False Rejection Rate per session when no updating.

According to this, the evolution of FRR and FAR through the time is represented in Fig. 1.

It is noticeable that the FRR values increase considerably over time, achieving even almost 70% of error. On the other hand, FAR remains constant since no updating process has been conducted so μ_8 of all the templates did not varied.

Therefore, it is deduced that users modify the manner they perform their identifying gesture significantly over time, so samples obtained when some time has elapsed could be quite different from those extracted when enrolling or at the first sessions. As a consequence, it seems evident that the error rates of the system would increase when it is used over a long period of time.

It is also highlighted that the behavior of the trend of FRR over time is, in general, increasing but continuous. Consequently, it might be deduced that users modify continuously but slightly between consecutive sessions the way they perform their gesture, thus, an updating strategy according to these characteristics could diminish error rates over time by updating the templates of the users in accordance with the modification of the performance of the gestures by the users.

4.2. Scenario 1: Always updating

In this scenario, the updating strategy proposed consists of substituting one of the samples of the template of each user by each original sample arriving the system. Therefore, this strategy represents an "always updating" scheme where the updating phase is carried out anytime the user performs his/her gesture to authenticate him/herself.

According to this, and due to the fact that the template is modified for each access, False Acceptance Rate is recalculated from each of the templates created through the study of all the database. As there are five samples for each gesture and each session, it is necessary to analyze all the falsified samples 95 times for each gesture (one for each variance of the template of the gesture). All these punctuations are stored separately by sessions in order to calculate the overall False Acceptance Rate, considering all the values, and the Daily False Acceptance Rates, regarding only the punctuations obtained in the session examined.

However, False Rejection Rate is calculated following the order of the samples obtained through different sessions, representing a real implementation of the system, where the original users try

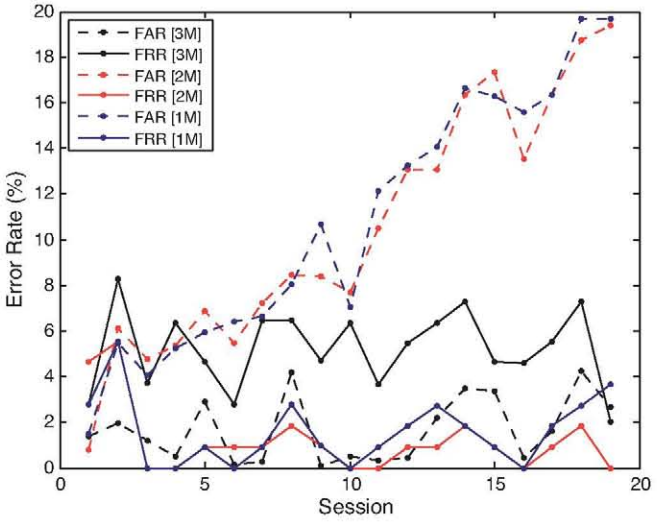

Fig. 2. Scenario 1: False Acceptance and False Rejection Rate per session.

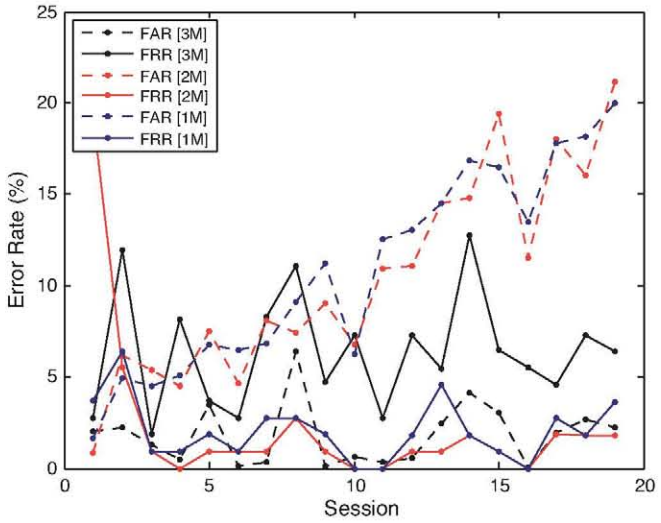

Fig. 3. Scenario 2: False Acceptance and False Rejection Rate per session.

to access five times each session and their templates are updated constantly. Therefore, the samples used to calculate Daily False Rejection Rates are those performed in the session of the session evaluated and compared with the corresponding template updated at that moment. Furtehrmore, the overall False Rejection Rate is calculated by considering the analysis of all the samples of all the sessions of each gesture.

Moreover, this scheme has been assessed when updating one of the three (3 M), one of the two (2 M) or always the same one (1 M) of the samples of the template of each user in a rotary manner. FAR and FRR on each session when updating following this scenario are presented in Fig. 2. In this figure, results when updating rotarily one of the three samples of the template are in black [3 M], whereas red lines stand for when the updating is performed within two samples [2 M] and blue lines when the sample of the template modified is always the same [1 M].

In Fig. 2 it can be observed that error rates get very high when not modifying all the samples of the template rotarily. However, in the [3 M] case, a very interesting result is obtained as error rates are moderately low and stable.

In addition to this, Table 2 provides the resulting values of the average and standard deviation of the FAR, the FRR and the average of both error rates over time obtained when updating one of three, of two or always the same of the template.

The results presented in Table 2 in conjuction with Fig. 2 mean that by following this strategy, an updating method is obtained able to adapt to how the users modify the manner in which they perform their signature in the air over time with a low average and stable error.

In conclusion, the strategy proposed in this scenario introduces great improvements as regards when no updating, reducing the average FAR to 1.67% and the FRR to 5.32% in a very constant manner (1.57% of standard deviation of the average of error rates) over time when the [3 M] case is selected.

4.3. Scenario 2: One-out-of-two updating

The second updating strategy considers updating the template often but not always. In particular, it is based on substituting one of the samples of the template any two accesses found, reducing the updating speed of the previous scenario by half.

According to this, in this article, template updating has been carried out in samples number one, three and five of each of the sessions of each gesture of the database (except session 1). Consequently, each user template was modified 57 times (three times for each session), and again, falsifying samples were analyzed for each variance of the template.

Therefore, following this scheme the evolution of the FAR and FRR are presented in Fig. 3.

In this figure, the same not recommended behaviours of cases [2 M] and [1 M] can be observed, since when several sessions have taken place, the FAR increases significantly. On the other hand, the [3 M] case provides a much better result, similar than in Scenario 1 but with slightly less stable.

Table 3 compiles the FAR and FRR averages and standard deviations over time when updating one of three, of two or always the same of the template.

Therefore, the [3 M] case in this scenario also provides good results in terms of performance of the FAR and FRR, but some more peaks have been found when analising these error rates which produce a high standard deviation value (2.40%) in the average of FAR and FRR.

4.4. Scenario 3: One-out-of-three updating

The third scenario proposes a very similar strategy to scenarios 1 and 2 but updating at a lower speed. In this scheme, only two updates were performed for each gesture and each session, corresponding to samples number two and four of the samples of each gesture and each session in the database.

Table 2
Average and deviation error rates for Scenario 1.

Case	FAR %	FRR %	(FAR + FRR)/2%
[3 M]	1.67 ± 1.42	5.32 ± 1.72	3.50 ± 1.57
[2 M]	10.50 ± 5.38	1.24 ± 1.50	5.87 ± 3.44
[1 M]	10.79 ± 5.61	1.53 ± 1.48	6.16 ± 3.55

Table 3
Average and deviation error rates for Scenario 2.

Case	FAR %	FRR %	(FAR + FRR)/2%
[3 M]	1.78 ± 1.67	6.37 ± 3.14	4.08 ± 2.40
[2 M]	10.54 ± 5.61	2.22 ± 4.35	6.38 ± 4.98
[1 M]	10.88 ± 5.53	2.07 ± 1.66	6.48 ± 3.60

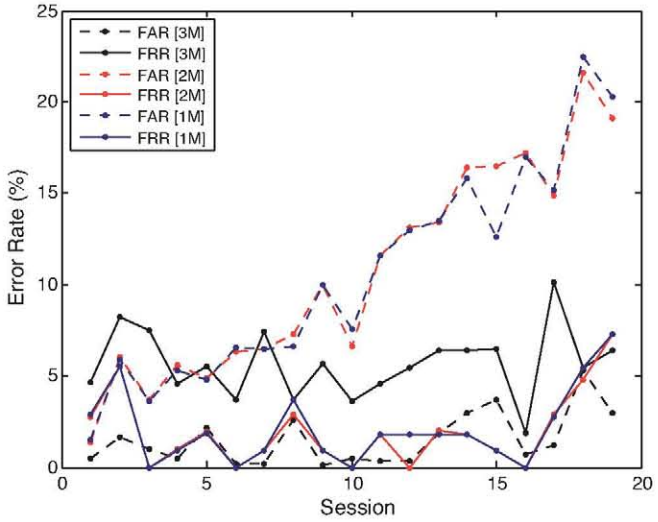

Fig. 4. Scenario 3: False Acceptance and False Rejection Rate per session.

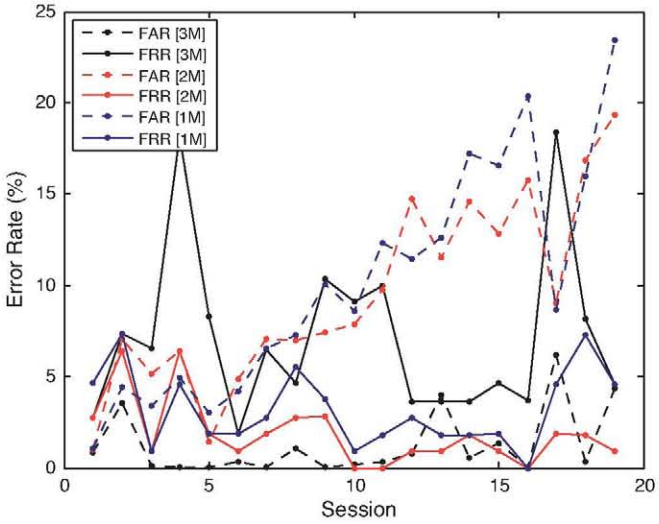

Fig. 5. Scenario 4: False Acceptance and False Rejection Rate per session.

This strategy means calculating 38 times the analysis of falsifying samples for each gesture (one for each template modified).

Following this scheme, the FAR and FRR obtained in each session are presented in Fig. 4, representing the [3 M], [2 M] and [1 M] cases.

Moreover, Table 4 presents the FAR and FRR averages and standard deviations over time when the strategy evaluated in this scenario is considerd. The results are detailed for all the cases assessed: when updating one of three, of two or always the same of the template.

The results of this strategy are quite similar but slightly worse than the previous two scenarios.

4.5. Scenario 4: In-1-session updating

The forth updating strategy proposed decreases the updating interval to once in each session. Therefore, the user's template will be updated on the first authentic access in the session. This means that the first access in the session is analyzed in comparison with the template of the previous session, and after that, if successful, the updating method is carried out. Therefore, the template is modified slowly enough to keep some variance between the gestures that make up the template.

This scenario has been also evaluated considering the different number of samples involved at updating, one of three (3 M), one of two (2 M) or only one of them (1 M).

Following this scheme, in Fig. 5 the results of FAR and FRR in each session for each number of samples involved are presented. In this figure it can be seen that although the FRR on average has a reasonable value, there are two peaks of error of almost 20% corresponding to two different sessions. Therefore, the updating interval should be lower than the one in this scenario.

The FAR and FRR averages and standard deviations over time following the strategy in this scenario are presented in Table 5.

The aforementioned peaks of error introduced a high deviation in case [3 M], which is the one which provides lower (FAR + FRR)/2 results.

Table 4
Average and deviation error rates for Scenario 3.

Case	FAR %	FRR %	(FAR + FRR)/2%
[3 M]	1.52 ± 1.45	5.69 ± 1.90	3.59 ± 1.67
[2 M]	10.77 ± 5.84	1.9910 ± 2.03	6.38 ± 3.93
[1 M]	10.55 ± 5.85	2.13 ± 2.06	6.34 ± 3.95

Table 5
Average and deviation error rates for Scenario 4.

Case	FAR %	FRR %	(FAR + FRR)/2%
[3 M]	1.19 ± 1.83	7.15 ± 4.64	4.17 ± 3.26
[2 M]	9.7718 ± 5.11	1.88 ± 1.81	5.82 ± 3.46
[1 M]	10.27 ± 6.34	3.18 ± 2.10	6.73 ± 4.21

4.6. Scenario 5: In-2-session updating

The fifth updating strategy proposes to update the template at a slower speed, at the beginning of each two sessions.

According to this scheme, Fig. 6 presents the general EER obtained when one of three, two or one of the samples of the template has been modified. Following this schema, a low average of FAR and FRR is obtained, but both rates appear to be very unstable. There are two peaks of almost 10% of error. Note in this figure that the FAR values are repeated in groups of two, since in the sessions where there is no updating process, the FAR does not change.

The FAR and FRR averages and standard deviations over time considering the strategy in this scenario are compiled in Table 6,

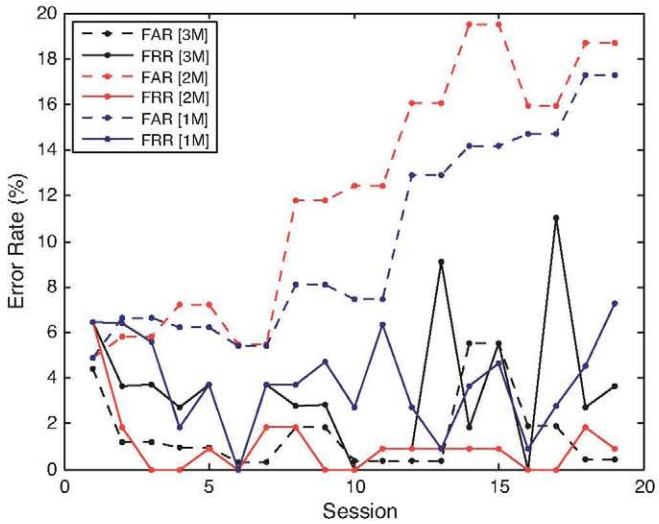

Fig. 6. Scenario 5: False Acceptance and False Rejection Rate per session.

Table 6
Average and deviation error rates for Scenario 5.

Case	FAR %	FRR %	(FAR + FRR)/2%
[3 M]	1.44 ± 1.71	3.43 ± 2.94	2.43 ± 2.32
[2 M]	12.53 ± 5.39	1.06 ± 1.48	6.80 ± 3.44
[1 M]	10.32 ± 4.35	3.81 ± 2.06	7.06 ± 3.21

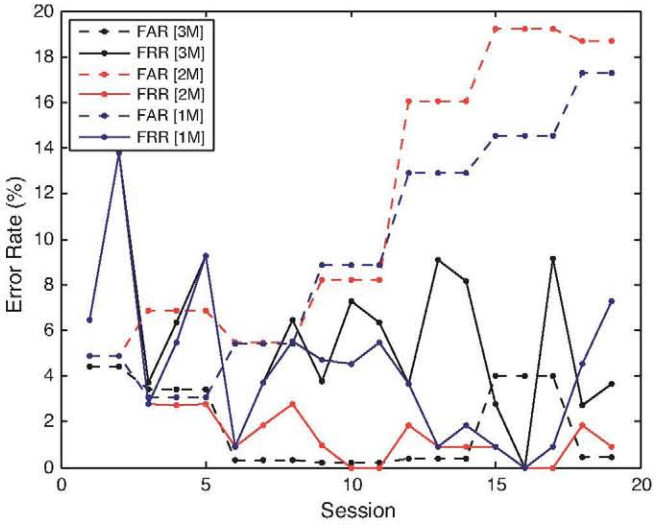

Fig. 7. Scenario 6: False Acceptance and False Rejection Rates per session.

where the results for [3 M], [2 M] and [1 M] are presented separately:

This scenario in case [3 M] provides a lower average error rate than in Scenario 1 (2.43% versus 3.50%), but it is much more unstable (2.32% versus 1.57%). This means that on average, it would not be so important to update at a high speed but, if the updating interval is long enough, some unexpected peaks of errors may appear, which might make the system unpredictable.

4.7. Scenario 6: In-3-session updating

Finally, the updating strategy evaluated in this scenario proposes a very low speed updating scheme. It suggest updating the templates at the beginning of each three sessions. Therefore, the templates are slightly corrected in respect to the templates obtained at enrollment.

Following this updating scheme obtains the FAR and FRR values over time represented in Fig. 7. In this updating strategy, the FRR per session is quite unstable, including a great number of high peaks.

In addition to this, Table 7 provides the values of both the average and standard deviation of the FAR, the FRR and the average of both error rates over time obtained when updating one of three, of two or always the same of the template.

The best average FAR and FRR obtained in this scenario (case [3 M]) is 3.55%, with a considerable amount of dispersion. Therefore, this updating strategy does not improve the results of previous scenarios.

Table 7
Average and deviation error rates for Scenario 6.

Case	FAR %	FRR %	(FAR + FRR)/2%
[3 M]	1.45 ± 1.83	5.65 ± 3.37	3.55 ± 2.60
[2 M]	12.41 ± 5.95	2.22 ± 3.18	7.32 ± 4.57
[1 M]	10.33 ± 5.00	4.34 ± 3.38	7.34 ± 4.19

5. Conclusions and future work

In this article a biometric technique based on the performance of identifying gestures while holding a mobile phone on a hand has been studied. When a gesture is performed, the accelerometers embedded in the mobile device obtain, at a sampling rate of 50 Hz, the values of the accelerations on each of the three axes of the space corresponding to the gesture performed.

For this purpose, a mathematical method to analyze difference between acceleration signals has been developed. This method is related to dynamic programming and global sequence alignment techniques. As a consequence of the application of this method to compare acceleration signals, eight different score definitions have been proposed to quantify the differences between performances of gestures.

Moreover, all these scores have been evaluated by analyzing a database of 80 users who have performed their identifying gesture (chosen by them) in front of a video camera. From these records, three different imposters attempted to forge each gesture.

According to this, from the study of the samples of 40 users, the Equal Error Rates and the optimal thresholds for each score have been obtained. From these optimal thresholds, False Acceptance and False Rejection Rates have been calculated with the remaining 40 users. For each score, 200 different configurations of the parameters h and σ have been tested. Each configuration of the algorithm for each score has been implemented 10 times with the random division of the users of the database used to train (EER and optimal threshold) and evaluate (FAR and FRR) in order to make the results the most independent possible from the users in the database.

In these conditions, the optimal score means to align the signals in comparison, interpolate the zeros included in the alignment process, calculate the sum of differences between the aligned signals in absolute value, and finally, normalize this value by the average of the two length of the original signals. Following this procedure, an EER of 3.42% has been obtained. With the threshold deduced from the obtention of EER, a FAR of 1.54% and FRR of 3.67% has been obtained when applying the algorithm to the samples of the users not used to calculate EER. Hence, this configuration (score, parameters h and σ and threshold) has been considered as the optimal and adopted for the rest of experiments.

Next, a temporal study of the technique has been introduced by studying the behavior of the system when users repeated their gestures over a long period of time. In particular, a database of 22 users repeating their gestures in 20 sessions spread over 4 months has been assessed.

As a consequence of this study, in Scenario 0, it has been concluded that a template updating strategy is necessary to make this biometric technique useful, since without it, errors increase considerably at users unconsciously modify the way in which they perform their gesture when they repeat it at different times. Actually, FRR achieves values of 70% of error when a long time has elapsed and no updating process has been carried out.

Consequently, six different updating methods were proposed, including three different updating implementation cases for each of them. From this study the following deductions are achieved:

- By alternatively modifying one of the three samples of the template, a much better performance is obtained than when one or two samples of the template remains without alteration.
- The more the template is updated the better and more stable performance is obtained.
- Even though the template is updated in several sessions, the average FAR and FRR results imply a good performance. However, it might appear some peaks of errors might appear whether the template was not updated for a long period of time.

Therefore, the optimal updating strategy studied in this article consists of substituting one of the three samples (the oldest one) of the template whenever an authentic sample is found.

Following this updating strategy, the average FAR and FRR results are 1.67% and 5.32%, respectively, which implies and average error of 3.50%. This average error is obtained with a standard deviation of 1.57%, the lowest in the article since it belongs to the most stable strategy. These results offer a huge improvement in comparison with when no updating process is implemented.

From the idea of "the best updating method is updated as often as possible", a very interesting future work to continue this updating approach would consist of obtaining a big database of people trying to enter the system at different intervals (several times a day, once a day, once each two days, once each weak, etc.), and studying different models to adapt the updating strategy to how often the users make their identifying gesture.

In summary, from this article it is concluded that it is possible to include some biometric security level in mobile applications where a personal authentication is required. In this article, this user authentication procedure is achieved through a biometric technique based on performing an identifying gesture created by users holding a mobile telephone with an embedded accelerometer in their hand. This biometric technique provides acceptable results when the optimal score presented in this article is selected as well as an optimal updating strategy is introduced.

References

Amayeh, G., Bebis, G., Nicolescu, M., 2009. Improving hand-based verification through online finger template update based on fused confidences. In: BTAS'09. IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems , pp. 1–6, iD: 1.

Chirillo, J., Blaul, S., 2003. Implementing Biometric Security, 1st edition. Hungry Minds, Incorporated.

Clarke, N., Furnell, S., 2007. Authenticating mobile phone users using keystroke analysis. International Journal of Information Security 6, 1–14.

de Santos Sierra, A., Avila, C., Vera, V., 2008. A fuzzy dna-based algorithm for identification and authentication in an Iris detection system. In: ICCST 2008: 42nd Annual IEEE International Carnahan Conference on Security Technology , pp. 226–232.

Durbin, R., Eddy, S., Krogh, A., Mitchison, G., 2006. Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids, 11th edition. Cambridge University Press.

Freni, B., Marcialis, G., Roli, F., 2008. Online and offline fingerprint template update using minutiae: an experimental comparison. Lecture Notes in Computer Science 5098, 448, SP: 441.

Guerra-Casanova, J., Sánchez-Ávila, C., de Santos-Sierra, A., del Pozo, G.B., Jara-Vera, V., 2010. A real-time in-air signature biometric technique using a mobile device embedding an accelerometer. In: Zavoral, F., Yaghob, J., Pichappan, P., El-Qawasmeh, E. (Eds.), NDT (1), Vol. 87 of Communications in Computer and Information Science. Springer, pp. 497–503.

ho Cho, D., Park, K.R., Rhee, D.W., Kim, Y., Yang, J., 2006. Pupil and iris localization for iris recognition in mobile phones. International Workshop on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, International Conference on & Self-Assembling Wireless Networks, 197–201.

Ijiri, Y., Sakuragi, M., Lao, S., 2006. Security management for mobile devices by face recognition. In: 7th International Conference on Mobile Data Management , pp. 49–149.

Jain, A.K., Flynn, P., Ross, A.A., 2007. Handbook of Biometrics. Springer-Verlag, New York, Inc., Secaucus, NJ, USA.

Jeong, D., Park, H.-A., Park, K., Kim, J., 2005. Iris recognition in mobile phone based on adaptive gabor filter. In: Zhang, D., Jain, A. (Eds.), Advances in Biometrics, Vol. 3832 of Lecture Notes in Computer Science. Springer, Berlin Heidelberg, pp. 457–463.

Kela, J., KorpipŠŠ, P., MŠntyjŠrvi, J., Kallio, S., Savino, G., Jozzo, L., Marca, S., 2006. Accelerometer-based gesture control for a design environment. Personal and Ubiquitous Computing 10, 285–299.

Li, Y., Yin, J., Zhu, E., Hu, C., Chen, H., 2008. Score based biometric template selection and update. In: FGCN'08. Second International Conference on Future Generation Communication and Networking, Vol. 3 , pp. 35–40, iD: 1.

Mäntyjärvi, J., Kela, J., Korpipää, P., Kallio, S., 2004. Enabling fast and effortless customisation in accelerometer based gesture interaction. In: Proceedings of the 3rd International Conference on Mobile and Ubiquitous Multimedia, MUM'04 ,. ACM, New York, NY, USA, pp. 25–31.

Marcialis, G.L., Rattani, A., Roli, F., 2008. Biometric template update: an experimental investigation on the relationship between update errors and performance degradation in face verification. Lecture Notes in Computer Science, SP: 684.

Rattani, A., Marcialis, G.L., Roli, F., 2008. Biometric template update using the graph mincut algorithm: a case study in face verification. BSYM'08: Biometrics Symposium, 23–28, iD: 1.

Rattani, A., Freni, B., Marcialis, G.L., Roli, F., 2009. Template update methods in adaptive biometric systems: a critical review. Lecture Notes in Computer Science, SP: 847.

Shabeer, H.A., Suganthi, P., 2007. Mobile phones security using biometrics. International Conference on Computational Intelligence and Multimedia Applications 4, 270–274.

Singh, R., Vatsa, M., Ross, A., Noore, A., 2009. Online learning in biometrics: a case study in face classifier update. In: BTAS'09. IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems , pp. 1–6, iD: 1.

Solami, E.A., Boyd, C., Clark, A., Islam, A.K., 2010. Continuous biometric authentication: can it be more practical? In: 12th IEEE International Conference on High Performance Computing and Communications (HPCC) , pp. 647–652, iD: 1.

Tao, Q., Veldhuis, R., 2006. Biometric authentication for a mobile personal device. Annual International Conference on Mobile and Ubiquitous Systems, 1–3.

Wayman, J.L., Jain, A.K., Maltoni, D., Maio, D., 2004. Biometric Systems: Technology Design and Performance Evaluation. Springer-Verlag, New York, Inc., Secaucus, NJ, USA.

Javier Guerra Casanova, was graduated as Telecommunications Engineer by Universidad Politécnica de Madrid in 2008. He is currently working at the Research Group in Biometrics, Biosignals and Security (GB2S) of the Universidad Politécnica de Madrid, as R&D engineer. He is currently a PhD student in ETSIT (Escuela Técnica Superior de Ingenieros de Telecomunicación). His PhD studies are focused on new biometric techniques based on behavioural characteristicss applied to mobile devices.

Dr. Carmen Sánchez-Ávila obtained her PhD in Mathematical Sciences in 1993, by the Universidad Politécnica de Madrid (UPM), being currently Associate Professor at UPM. She is now in leadership of the Research Group in Biometrics, Biosignals and Security (GB2S) of UPM, involved in project research and development concerning a broad range of applications, from Mobile Security Services based on Biometric till Fast Cryptographic Protocols, Secure Transmission of Large Packages of Data and Crypto-Biometric. She is an expert in Biometrics and Cryptography Security and member of SC37 Standardization Committee.

Alberto de Santos Sierra received the degree of Telecommunication Engineer in Escuela Técnica Superior de Telecomunicación (ETSIT), finishing his studies in Vrije Universiteit Amsterdam with a Final Master Project based on Iris Recognition, in July, 2007. At present, he is working at the G2BS in topics related to Biometric Recognition Systems, Stress Biometry and Crypto-Biometric. He is currently a PhD Student in ETSIT focusing on Biometrics based on Hand Recognition and other physiological characteristics oriented to mobile devices.

Gonzalo Bailador received his PhD degree in Computer Science from University Politécnica de Madrid. He is currently working in the research group GB2S focused on applying pattern recognition techniques to the analysis of temporal signals and mass spectrometry data. His research interests include gesture recognition, gait recognition, odour identification and robotics.