Performance Evaluation Of Behavioral Biometric Systems

Fouad Cherifi, Baptiste Hemery, Romain Giot, Marc Pasquet, Christophe Rosenberger

Performance Evaluation Of Behavioral Biometric Systems

F. Cherifi, B. Hemery, R. Giot, M. Pasquet, C. Rosenberger

G R E Y C L a b o r a t o r y
E N S I C A E N – C N R S – U n i v e r s i t y o f C a e n , F r a n c e

A B S T R A C T

We present in this chapter an overview of techniques for the performance evaluation of behavioral biometric systems. The BioAPI standard that defines the architecture of a biometric system is presented in the first part of the chapter... The general methodology for the evaluation of biometric systems is given including statistical metrics, definition of benchmark databases and subjective evaluation. These considerations rely with the ISO/IEC19795-1 standard describing the biometric performance testing and reporting. The specificity of behavioral biometric systems is detailed in the second part of the chapter in order to define some additional constraints for their evaluation. This chapter is dedicated to researchers and engineers who need to quantify the performance of such biometric systems.

K E Y W O R D S

Systems Evaluation, Biometrics, Authentication, Metrics, Behavioral biometrics

I N T R O D U C T I O N

Biometrics is now a technology that is present in our daily life. It is used as for example in airports (passport verification), offices (access control, biometric USB key...) and even in some places in the world for banking operations... Different biometric modalities can be used for the identification / verification of an individual (face recognition, keystroke dynamics recognition, DNA analysis...) (Mahier et al., 2008).

The characterization of an human by considering its behavior in its daily life operations (gait, signature dynamics, voice...) (Han et al. 2006; Muramatsu & Matsumoto, 2007; Petrovska-Delacretaz et al., 2007) or through its interactions with a computer (mouse dynamics, keystroke dynamics...) represents an interesting and open area in research (Hwang et al., 2006; Orozco et al., 2006).

The performance evaluation of such biometric systems is very important for many reasons:

- To be used in a real (that is to say in an industrial) context, the quality of a biometric system must be precisely quantified. The context of use, the efficiency, the robustness of the algorithm must be defined to determine if it fulfills the requirements of a particular industrial application (logical access, physical access, e-commerce...) ;

- The comparison of different biometric modalities is essential to qualify their relative advantages and drawbacks ;

- The performance evaluation is also necessary in order to facilitate the research in this field (Hemery et al., 2006). We need a reliable evaluation method in order to put into obviousness the benefit of a new biometric system.

The objective of this chapter is to make an overview on evaluation techniques that are used in the state of the art to quantify the performance of behavioral biometric systems. An engineer or a researcher will find in the proposed chapter, the different criteria or methods he can use to validate a biometric system he intends to use in a real context. A behavioral biometric system can be evaluated by considering the general approach to evaluate a biometric system while taking into account the specificity of this type of modality.

The plan of the chapter is given below. In the section 1, we present the general approaches for the evaluation of a biometric system. It necessitates generally to use a benchmark database (Hemery et al., 2007) and a set of criteria (computation time, FAR...). The benchmark database can be composed of real biometric templates or synthetic ones. We present different solutions from the state of the art. Section 3 focuses on specificities of behavioral biometric systems. We present their specificities that must be taken into account for their evaluation. Section 4 concerns the future trends that must be achieved in order to facilitate research progress in this domain. We conclude this chapter in section 5.

GENERAL EVALUATION METHODOLOGIES

1. Introduction

A biometric system is composed of different steps (see Figure 1). There are mainly two processes in the use of a biometric system. The enrollment phase has for objective to determine a model of an individual given the characteristics acquired by the selected biometric sensor. The identification / verification phase uses this model to make a decision an individual.

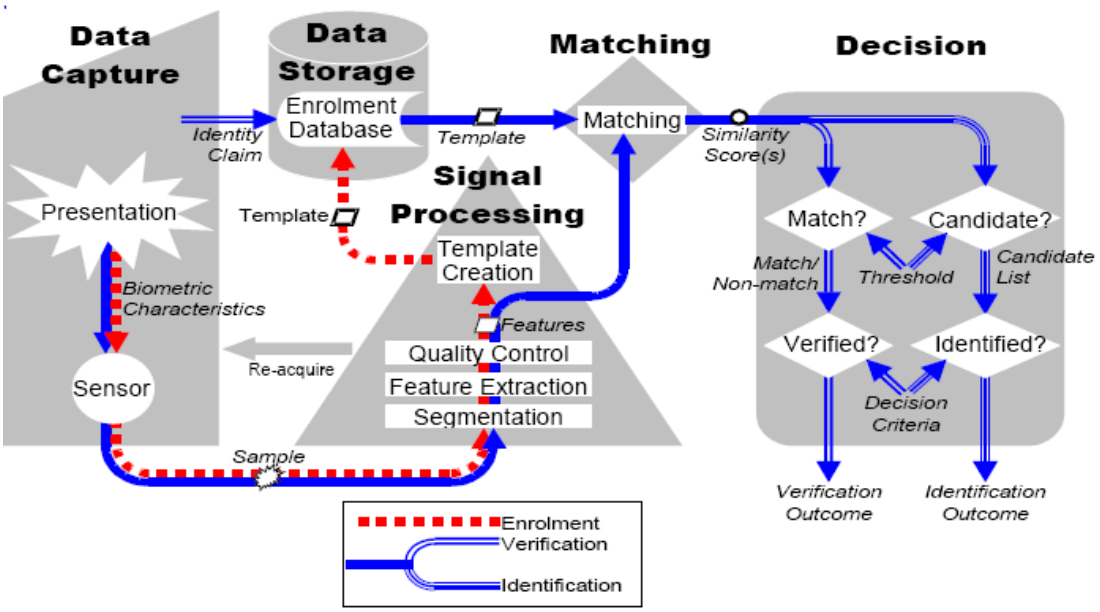


Figure 1: Diagram summarizing the various phases of a biometric system (Source: ISO/IEC JTC 1/SC 37 Part 1 Overview Standards Harmonization Document)

The international standards committee for biometrics within ISO (ISO/IEC JTC1 SC37) developed a complete specification and reference implementation for a standardized API (BioAPI Consortium, 2005). The purpose of the BioAPI Specification is to define an open system standard application program interface (API) which allows software applications to communicate with a broad range of biometric technologies in a common way.

Figure 2 shows the interaction between the three BioAPI 2.0 components: applications, BioAPI Framework, and BSPs (Biometric Service Providers). The BioAPI 2.0 specification implements two APIs. The first one is the API which is the interface between the BioAPI Framework which supports the functions in the API specification and application. The second is the Service Provider Interface (SPI) that is the interface between the BioAPI Framework which invokes the functions in the SPI specification and to support the functions of the SPI specification and BSPs. The BioAPI Framework is responsible for the management of BSPs and for the mapping of function calls from an API function to an SPI function within the appropriate BSP.

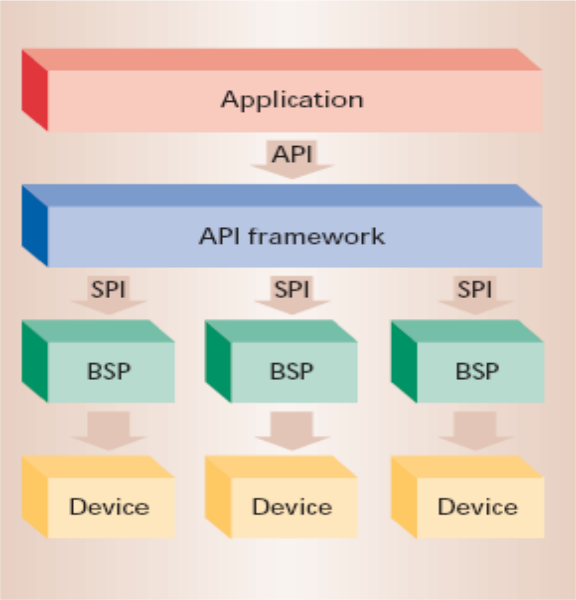


Figure 2: BioAPI components (source BioApi, 2005).

The performance evaluation of biometric systems is a crucial problem. It is generally realized within three contexts (ISO, 2006):

- Technology evaluation: It consists in testing an algorithm on a standardized corpus. The objective is to determine if the developed biometric system *a priori* meets the requirements. Testing is carried out using offline processing of the data and the results of technology tests are repeatable ;

- Scenario evaluation: The testing is carried out on a complete system in an environment that models a real-world target application of interest. Test results will be repeatable only to the extent that the modeled scenario can be carefully controlled ;

- Operational evaluation: It is an online test in real conditions. In general, operational test results will not be repeatable because of unknown and undocumented differences between operational environments.

The performance evaluation of a biometric system generally considers the quality of the input data and the output result. In order to evaluate them, we generally use an empirical process by considering the system as a black box (Thacker et al., 2008). In this case, the internal working of the associated algorithms is not studied. The black box generates an output result given a biometric template as input and a set of parameters. We identified within this context different issues in the evaluation of biometric systems:

- Quality control of the biometric template: This quality is necessary to be quantified before the enrollment or verification/identification step ;

- Definition of benchmark databases: It is a challenge as it is used in the two first evaluation contexts ;

- Performance evaluation: The characterization of a biometric system uses many metrics such as error rates or average verification time ;

- Subjective evaluation: Many other aspects must be taken into account such as the user acceptability or its confidence.

We detail all these issues in the next sections.

2. Quality control of the biometric template

A biometric system is composed of two main components: a sensor that permits to acquire the biometric template and some algorithms for the enrollment and the verification/ identification steps. The quality of the biometric template is essential to guarantee a correct behavior of the biometric system. Many problems can alter this quality mainly because of three reasons (ISO, 2006):

- Problems due to the sensor: incorrect parameterization (volume for audio, focus for image based sensor...), dirty sensor (as for example, optical fingerprint sensor), transmission error...

- Problems due to the user: incorrect use of the sensor (too far from the microphone, not in the field of the camera...), behavior (stress, tension, mood or distractions), personal characteristic (accent, handicap...), personal modifications (haircut change, keystroke...)...

- Problems due to the environment: conditions of acquisition (noise, light, humidity...)...

In the BioAPI standard, the quality of the biometric template can be evaluated by the BSP. If the quality (a score between 0 and 100) is considered as insufficient, the user is asked to acquire again the template.

Many specifications by the ISO organization defined some evaluation criteria for the quality of few biometric templates such as face, fingerprint or Iris. Other biometric modalities are currently studied such as the signature, voice or hand shape. If we consider as for example the face modality, the evaluation of the template takes into account the resolution of the image, the size of the face in terms of pixels in the image or the compression rate used to store the face image.

3. Definition of benchmark databases

In order to compare different biometric systems, we need generally to compute their performance following the same protocol (acquisition conditions, test database,

metrics...). The testing database contains many samples specific to a biometric modality and each sample is associated to an individual. By comparing the performance of different systems on the same database and with the same experimental conditions (number of samples used in the enrollment step, thresholds), we can decide which system performs better. In this case, it provides us a relative evaluation of biometric systems.

These benchmark databases aim to be as close as possible as real use cases. By the way, a database must contain enough samples from an individual for the enrollment step. Moreover, a database is generally composed of two parts. The first one is used for the enrollment and the second one for the identification/verification task. A database must also contain a large number of individuals because the performance of biometric systems generally decreases as the number of user increases. Finally, the samples must represent most of different possible alterations that could be seen in a real use, as for example noisy or incomplete biometric data.

A benchmark database can contain real samples from individuals, which reflect the best the real use cases. Nevertheless, it is difficult to create such a database for several reasons. First of all, it can be difficult and costly to collect samples from a high number of individuals. Moreover, all samples must be acquired in the same conditions. This constraint can be very difficult to fulfill (as for example the guarantee to have the same lighting conditions for the face capture). Samples must then be acquired with some alterations to represent difficulties during the identification/verification task. Finally, each sample must be annotated by an human. A database can be specific to a modality, like the USF HumanID gait database (Sarkar et al., 2005), but can also be multimodal like the MCYT-100 database (Ortega-Garcia et al, 2003) which contains samples of fingerprint and signature for the same individual.

A benchmark can also contain synthetic samples. The creation of such a database is easier but is less significant. The main advantage of synthetic database is that alterations on samples are fully controlled. This enables to verify the robustness of a biometric system face to a specific alteration. Such a database has been realized for fingerprints (Cappelli et al., 2002) as for example and used in the Fingerprint Verification competition in 2006 (FVC 2006). Figure 3 shows some examples of synthetic fingerprints. Alterations are simulated to make the fingerprint more realistic.

Figure 3: Some examples of synthetic fingerprints generated by SfinGe (Maltoni, 2004).

4. Performance evaluation

The performance evaluation has for objective to provide some quantitative measures on the efficiency of biometric systems. The classical statistical metrics used to quantify the performance of a biometric system are:

- Computation time: the necessary time for the acquirement, enrollment, verification / identification ;
- True positive (TP): number of users that have been correctly authenticated ;
- False positive (FP): number of impostors that have been authenticated ;
- False reject rate (FRR): Proportion of authentic users that are incorrectly denied. It is calculated as:

  FRR = 1-TP/(number of genuine users)

- False accept rate (FAR): proportion of impostors that are accepted by the biometric system. It is calculated as:

  FAR = FP/(number of impostor users)

- Failure-to-enroll rate (FTE): proportion of the user population for whom the biometric system fails to capture or extract usable information from biometric sample. This failure may be caused due to behavioral or physical conditions pertaining to the subject which hinder its ability to present correctly the required biometric information ;
- Failure-to-acquire rate (FTA): proportion of verification or identification attempts for which a biometric system is unable to capture a sample or locate an image or signal of sufficient quality ;
- False match rate (FMR): The rate for incorrect positive matches by the matching algorithm for single template comparison attempts. FMR equals FAR when the biometric system uses one attempt by a user to match its own stored template ;
- False non-match rate (FNMR): The rate for incorrect negative matches by the matching algorithm for single template comparison attempts. FNMR equals FRR when the biometric system uses one attempt by a user to match its own stored template ;
- Identification rank: It is the smallest value *k* for which a user's correct identifier is in the top *k* identifiers returned by an identification system ;
- Receiver operating characteristic curve (ROC curve): The method most commonly used to assess the performance of a biometric system is the ROC curve. The aim is

to plot a curve representing FAR according to the FRR. In order to plot this type of curve, we have to changes the value of the decision threshold. For each value of the threshold, we calculate the associated FRR and FAR that we plot on the curve. The advantage of this method is that it gives a compact representation of the performance of a biometric system through a single curve allowing the comparison of different biometric systems. In order to compare easily several biometric systems, we can then compute the area under the curve AUC and the equal error rate ERR where FAR = FRR. The optimal result is obtained if the AUC equals 1 and the ERR equals 0 ;

- Detection error trade-off curve (DET curve): DET curve (Adler et al., 2007) is a ROC curve which has its linear scale replaced by a scale based on a normal distribution, to make it more readable and usable. In this case, the curve flattens and tends towards the right. The benefits of the DET curves are the same as those of ROC curves, but they allow in addition to compare biometric systems that have similar performance. An example of a DET curve can be seen on Figure 4 ;

- Cumulative match characteristic curve (CMC curve): This curve plots the identification rank values on the x-axis and the probability of correct identification at or below that rank on the y-axis.

- Precision/recall curve (PR curve): This curve (Muller et al, 2001) has a similar behavior to ROC curves. In order to draw the PR curve, we plot the positive predictive value (PPR = TP /(TP+FP)), also known as the precision versus the recall. We can then compute AUC and ERR in a similar way as in ROC curves. One advantage is that we do not need the number of true negative in this method.

The most used methods are the DET curves, ROC curves and the PR curves.

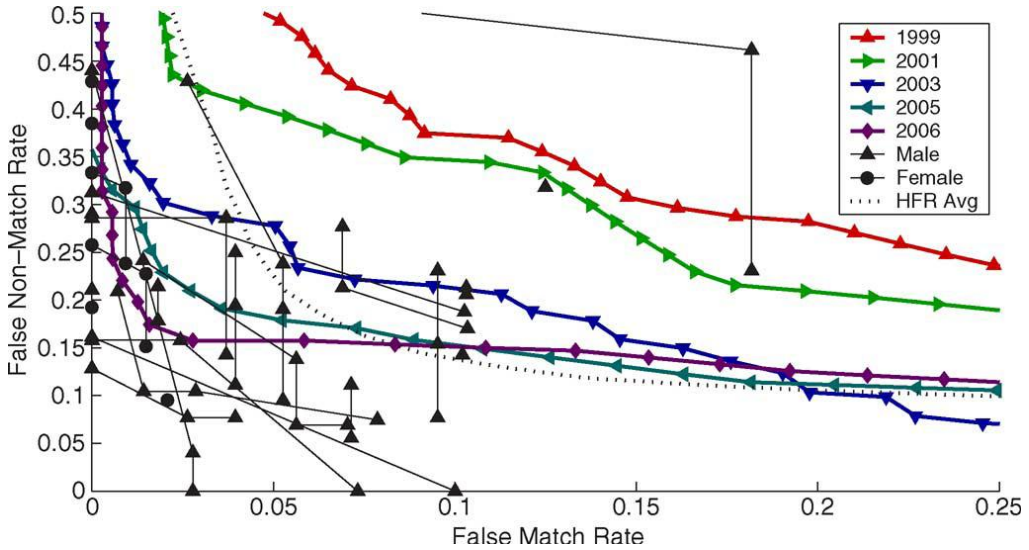


Figure 4: Example of DET curves used for the performance comparison of face recognition systems (source Adler et al., 2007).

5. Subjective evaluation

The performance evaluation is not the only thing to take into account when we have to consider a biometric system. To be accepted by users in real conditions (physical or logical access as for example), a biometric system must fulfill some other properties such as the acceptability, the easiness of use or the confidence in the system.

The acceptability denotes the way how users perceive the biometric system and interact with it. The acceptability is highly dependent of the culture of users. As for example, Asian users hardly accept to have a contact with a biometric sensor. This implies that biometric systems with contact less sensors, such as iris recognition, are better accepted by Asian users that biometric systems that need a contact, such as the fingerprint recognition. Another example, European users prefer fingerprint recognition to iris recognition.

The confidence in a biometric system is very close to its acceptability, as it is also highly dependent of the culture of users. It denotes how the reliability of a biometric system is perceived by users. Generally, users have a better confidence in biological biometric system than in behavioral biometric system. Fingerprint recognition or a DNA analysis is often considered, quite rightly, to be better than voice recognition. In the same time, the more the modality is efficient, such as the DNA analysis, the more it invades privacy and the less the acceptability is high.

The easiness of use depends on the quality of the sensor and the interface with users. It also depends on the time necessary for the identification: the system is not easy to use if several minutes are needed between the time the user gives his biometric data and the time the biometric system identifies the user. Another point that could be considered is the time necessary for the enrollment step and its easiness.

It is possible, especially during the operational evaluation, to ask the users to fill a form in order to have their opinion on these aspects. This permits to have complementary information on different biometric systems. A statistical analysis of the answers must be performed in order to keep only reliable users using the correlation factors or Chi square tests.

Finally, biometric systems are confronted to juridical problems concerning data corresponding to template of biometric system users. This depends mainly of the country where the biometric system is used. Each country has its own lawn concerning the protection of private data. As for example, in France, the use of computer devices to treat and save private data is regulated by the CNIL (French data protection authority). By the way, to use a biometric system, a French company must warn the CNIL and asks for their authorization before being able to collect samples and names used in biometric systems.

We detailed in this section the general scheme for the evaluation of any biometric system. We focus in the next section on behavioral ones and we put into obviousness their specificity.

THE SPECIFICITY OF BEHAVIORAL BIOMETRICS

Behavioral biometric systems are specific. Many characteristics make them difficult to define and to quantify their performance:

- The biometric template contains generally temporal information. As for example, for keystroke dynamics analysis, we generally use a template composed of a set N value couples {(Di,Fi) i=1..N} where N is the number of characters in the password, Di is the duration time the user presses a key and Fi is the time between this key and the next one in the password typing. For voice recognition systems, the biometric template is a sampled signal. Thus, the biometric template is generally quite important in size meaning that the parameters space is high ;

- The biometric template can change with time according to users. If we keep in mind the example of keystroke dynamics analysis, users with time learn how to type more efficiently their password. That means that the biometric template can be quite different compared to the one obtained after the enrollment step (Hocquet et al., 2007). Another example, the dynamics of signature can also change a lot with time as it becomes a reflex for the user. This variability has multiple consequences. The first one concerns the number of templates for the enrollment step that is generally higher than other types of biometric systems. The second consequence is that the verification / identification algorithm should take into account this variability in order to make a correct decision. Another point concerns the testing of such biometric systems with biometric data that must embed this difficulty ;

- The behavior as biometric characteristic can be very different for an individual given its age, culture and experience. The evaluation of a behavioral biometric system is often realized considering a large diversity of users.

We focus in the next sections on impacts of these remarks on the evaluation of behavioral biometric systems.

1. Benchmark definition

Benchmark definition is really important for the performance evaluation of biometric systems.

As mentioned previously in the chapter, a benchmark database can be composed of real biometric templates (from test users) or synthetic ones. The definition of synthetic templates is easier for behavioral biometric data. Indeed, many behavioral modalities can be synthesized rather easily such as keystroke dynamics, voice, lip movements, mouse dynamics, signature dynamics... For morphological biometric modalities, it is much more difficult to do. The ability to generate more easily synthetic biometric templates is an advantage for the evaluation of such systems.

Generally, a biometric model (generated after the enrollment phase) is computed for the same person given 2 or 3 capture sessions. As for example, the AR face database has been created considering two sessions with an interval between them of 15 days (Phillips & al., 2000). The difficulty of behavioral biometric systems is that the biometric template naturally changes with time. Indeed, a human is a nice machine who wants to do things quicker. As a consequence, a benchmark database for behavioral modalities needs more capture sessions in order to take into account this variation. As for example, Awad & Traore in the approach they proposed in 2005 for computer intrusion detection with behavioral biometrics has been validated with biometric data acquired during 9 sessions (Awad & Traore, 2005). The number of capture sessions is important but also the period of time between them. This shows the difficulty and the cost of such benchmark definition for this type of biometric modality.

The variability of behavioral biometric templates is really important if we compare morphological ones. Indeed, the fingerprint of individuals from different cultures or age is not so different. If we consider now the behavioral biometric modalities such as the keystroke dynamics, voice or gait, the associated template can be very different from individuals at different ages. As a consequence, the benchmark database must embed all the variability of biometric templates to be representative of real applications. As for example, Janakiraman & Sim (Janakiraman & Sim, 2007) tested their keystroke authentication method on users that were Chinese, Indian or European origin. This is so an additional constraint for the definition of benchmark databases.

2. Robustness analysis

The behavior of an individual is very dependent on many factors like his mood, emotion, tiredness or health... As for example, voice recognition systems are very sensitive to all these factors. In order to be used in a real context, one has to test the robustness of a biometric system face to all these modifications.

Behavioral biometric systems can be very sensitive according to the sensor (keystroke dynamics, mouse dynamics...). The behavior of an individual can be different for multiple sensors. As for example, the template generated during the enrollment based on keystroke dynamics for an individual and a given keyboard cannot easily be used for the verification on another keyboard (Clarke & Furnell, 2007). Indeed, the performance in term of EER can be quite high (>10%) in this case.

Another point concerns the robustness of behavioral biometric systems face to attacks. The main difficulty for these systems is that anybody can try to duplicate a biometric template. As for example, it is not very hard to launch the verification given its keystroke dynamics, voice or gait. That does not mean that the chance to be authenticated is necessary higher but it is very easy to make a try. For morphological biometric systems, it is much more difficult even if as for example, fingerprints can be duplicated with some effort to launch one verification.

3. Discussion

In order to evaluate a behavioral biometric system, one could use the general methodology described in the previous section by using benchmark databases and classical performance metrics. Nevertheless, several aspects must be taken into account to consider the specificity of such systems:

- The number of sessions for the definition of biometric templates for testing such a biometric system must be high (≥ 3) ;
- The sessions must be relatively spaced in order to take into account the natural change of behaviors of individuals ;
- Behavioral biometric templates can be in general easily synthesized. This approach for the definition of a benchmark database is interesting. It allows to test a large number of biometric templates and to control their alterations to quantify the robustness of the system ;
- The benchmark database must contain some fake biometric templates to also test the robustness of the system ;
- A benchmark database must embed a large diversity of users (culture, age...);
- The performance evaluation of behavioral biometric systems must be realized using the same sensor during the enrollment and verification / identification steps.

E X P E R I M E N T A L R E S U L T S

We present the evaluation of a keystroke dynamics verification system (Hocquet et al., 2007) as illustration of the proposed methodology. We first detail the protocol of the experiment we realized.

1. Protocol

We asked 15 individuals to participate for this experiment. Figure 5 shows some information on these individuals considering their age. Three females and twelve males participated to this experiment. All the involved individuals use a computer in their daily life. We explained them before starting the objectives of the experiment and the acquisition process. Each user tried two times in the first session before we record the biometric data.

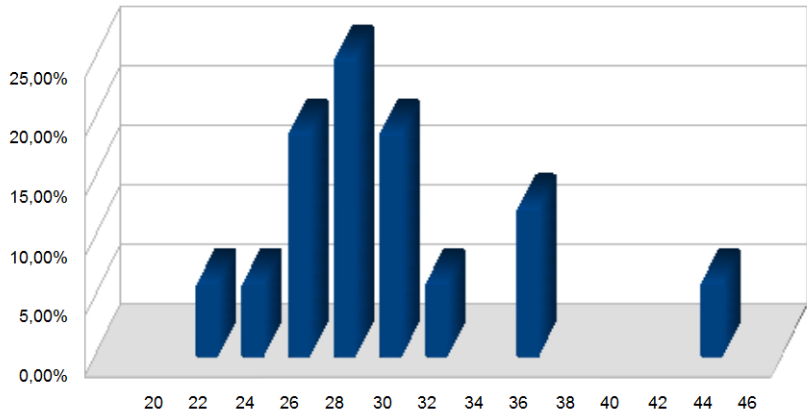


Figure 5: Repartition of users' ages for the experiment

Figure 6 presents the dates where sessions have been realized.

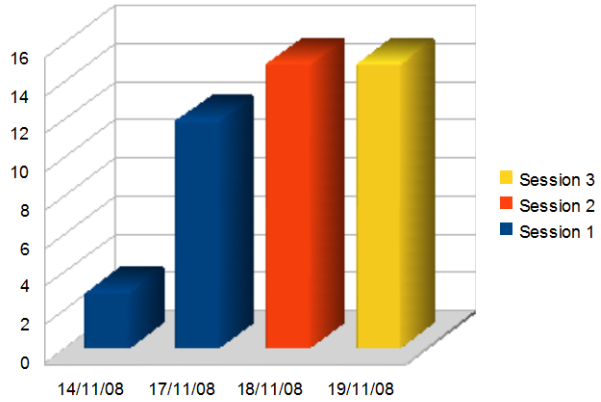


Figure 6: Session dates for the experiment

The biometric template contains: time between two keys pressure, time between two keys release, time between one pressure and one release and time between one release and one pressure. We asked each individual for each session to type 5 times the same password "GREYC Laboratory". We measured many data such as the time necessary for each user to type all of them, the number of mistyping and of course, the data that will be used as biometric template. To quantify the objective evaluation performance of this biometric system, we used for each individual, 5 biometric templates for the enrollment step and the 10 last for the verification one. To complete this objective evaluation experiment, we also realized a subjective evaluation test by asking the users to answer the following questions:

Q1: Is the verification fast? **Yes, no**
Q2: Is the system easy to use? **Yes, no**
Q3: Are you ready to use this system in the future? **Yes, no, do not know**
Q4: Do you feel confident in this system? **Yes, no**
Q5: Do you feel embarrassed when using this system? **Yes, no, do not know**
Q6: What is your general appreciation of the system? **Very good, good, average, bad**

2. Results

Figure 7 presents some statistics on the capture process. The FTA value is quite important for some individuals (the user wants to type the sentence too fast, users not enough concentrated…). The average FTA value equals 16% which is important.

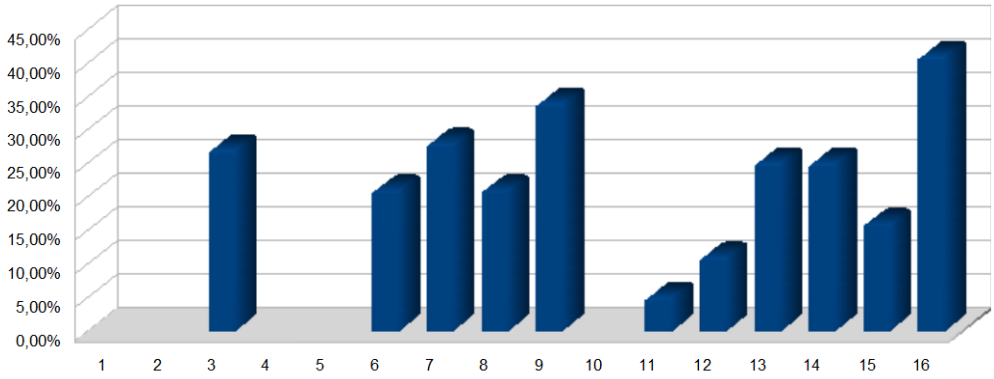


Figure 7: FTA value for each individual for all sessions

Figure 8 shows as illustration the biometric template acquired for an individual in a session. We can remark that the biometric template is quite stable.

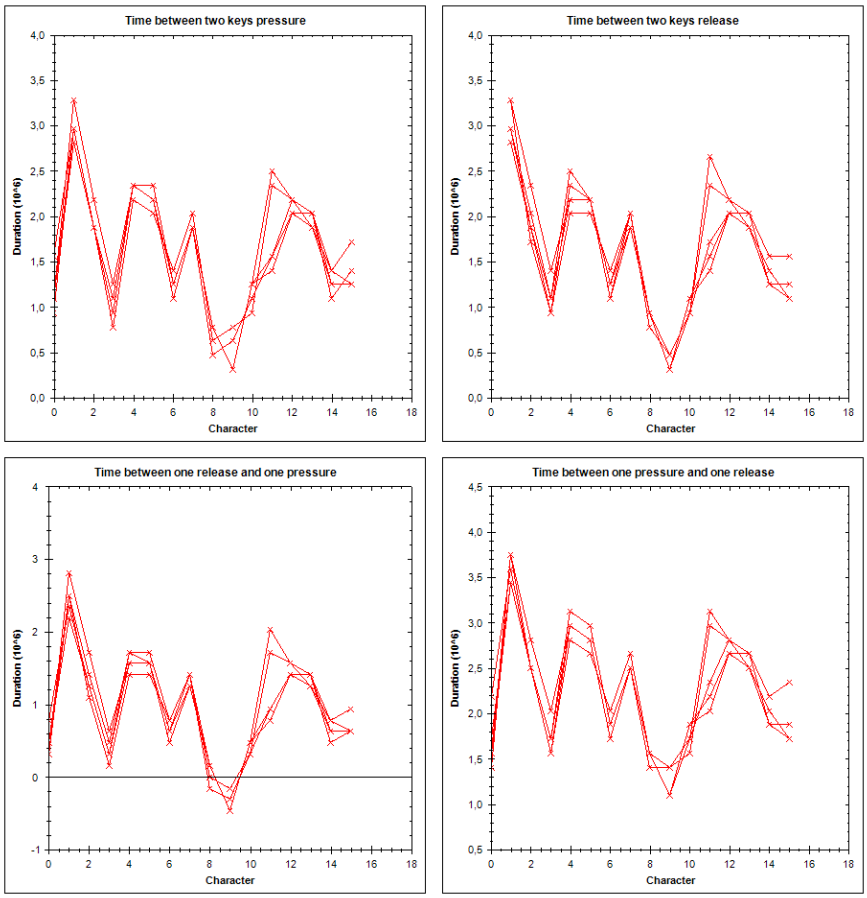
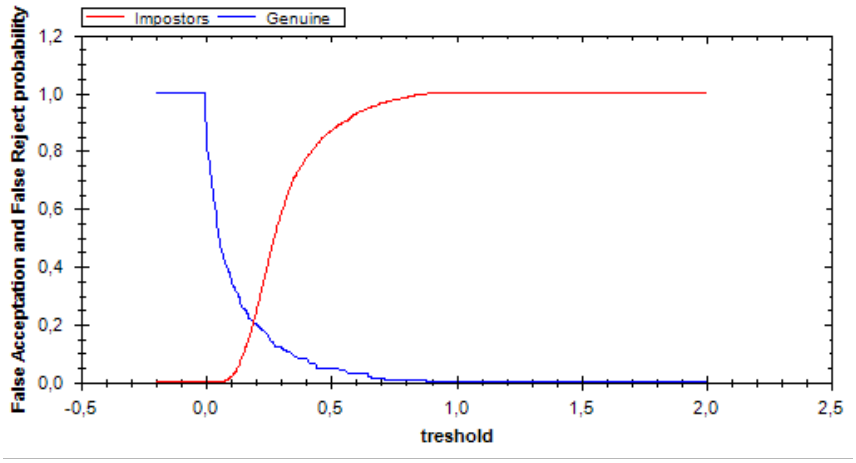
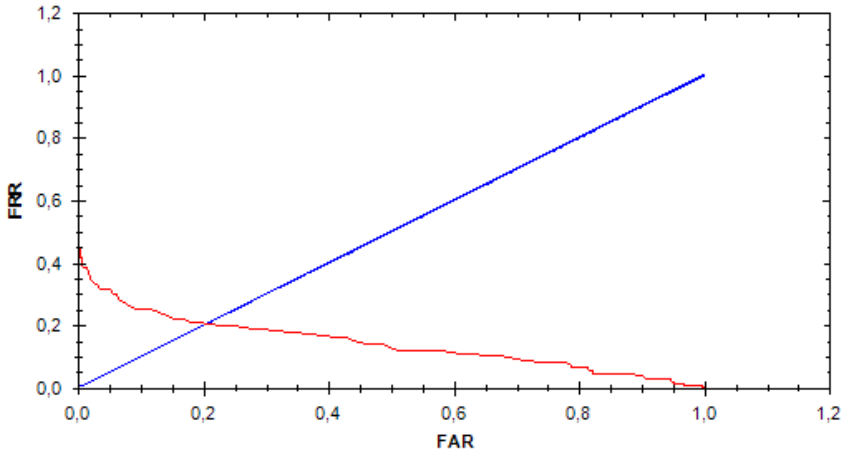


Figure 8: Plot of the biometric data for an individual for one session (5 acquisitions)

The resulting ROC curve is given in Figure 9. The computed EER value equals 20.5 %. The performance of this biometric system is not really important for this study as the main objective is to illustrate in this part, the different elements to take into account for its evaluation. If we want to obtain a FAR value equals to 0, the FRR value equals 50.16%. That means if we want none impostor, we have to set the value of the threshold having as consequence to reject genuine in 50.16% cases.



(a) Scores distribution



(b) Roc curve

Figure 9: (a) scores distribution and (b) Roc curve

In order to quantify the robustness of the selected biometric system, we made an experiment consisting in generating random synthetic keystroke dynamics. Given the 5 biometric templates used in the enrollment step, we compute an average biometric template denoted *E[T]*. We generated different biometric templates T_i *i=1:15* given *E[T]* and adding a random alteration by controlling its standard deviation. Figure 10 shows some examples of altered biometric templates given *E[T]* for an order value equals to 2.

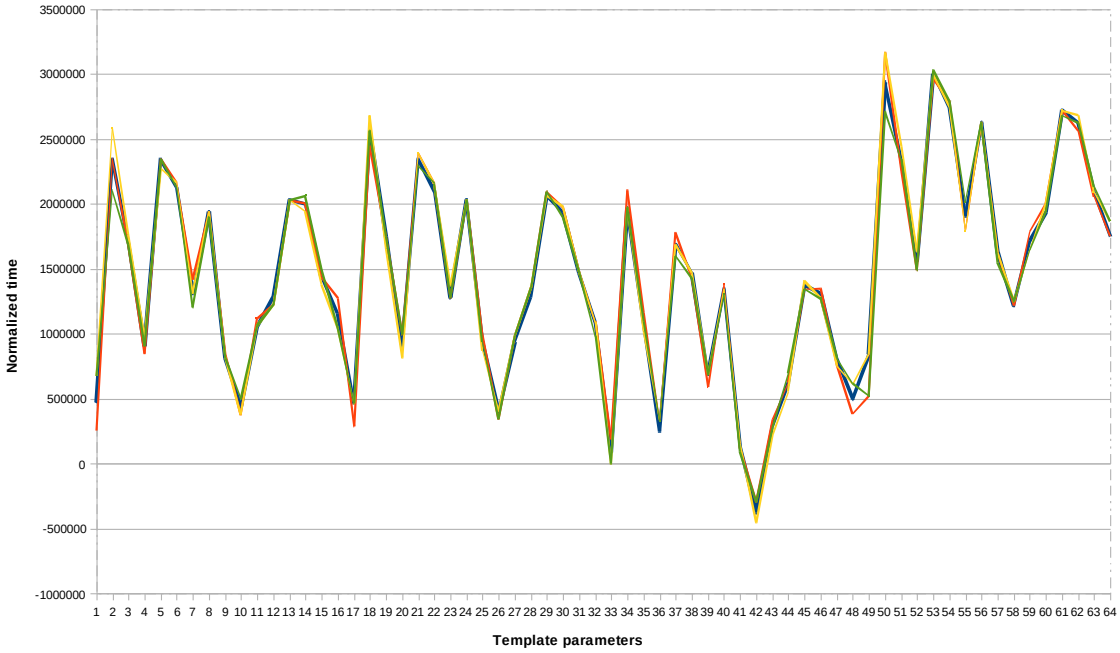


Figure 10: Three examples of random alteration (order 2) of the average biometric of an individual

Figure 11 shows the evolution of the EER value considering the standard deviation of the alteration of *E[T]*. We can notice that a small alteration has a great impact on the EER value; this means that this biometric system is not very robust.

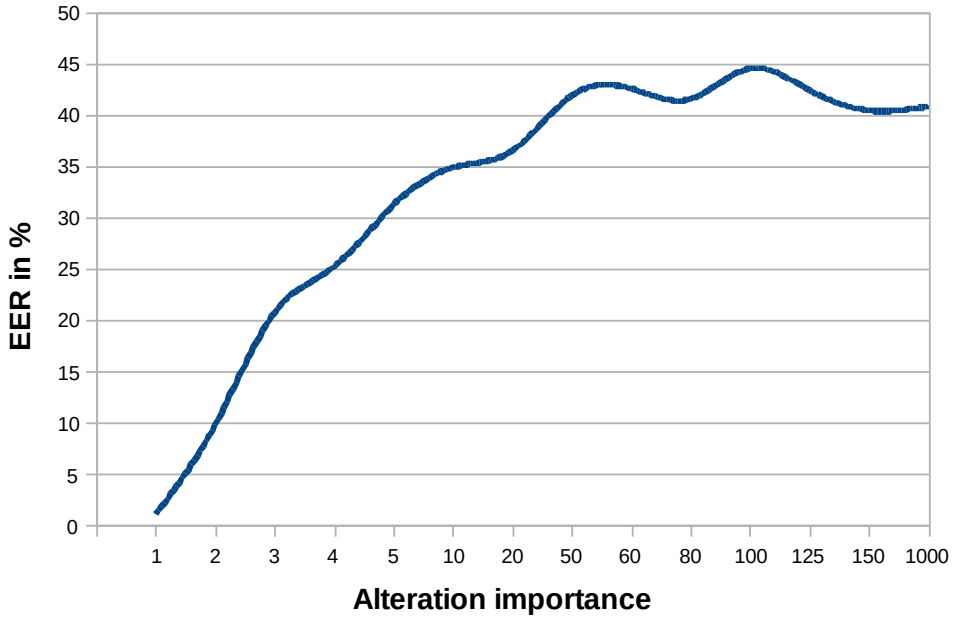


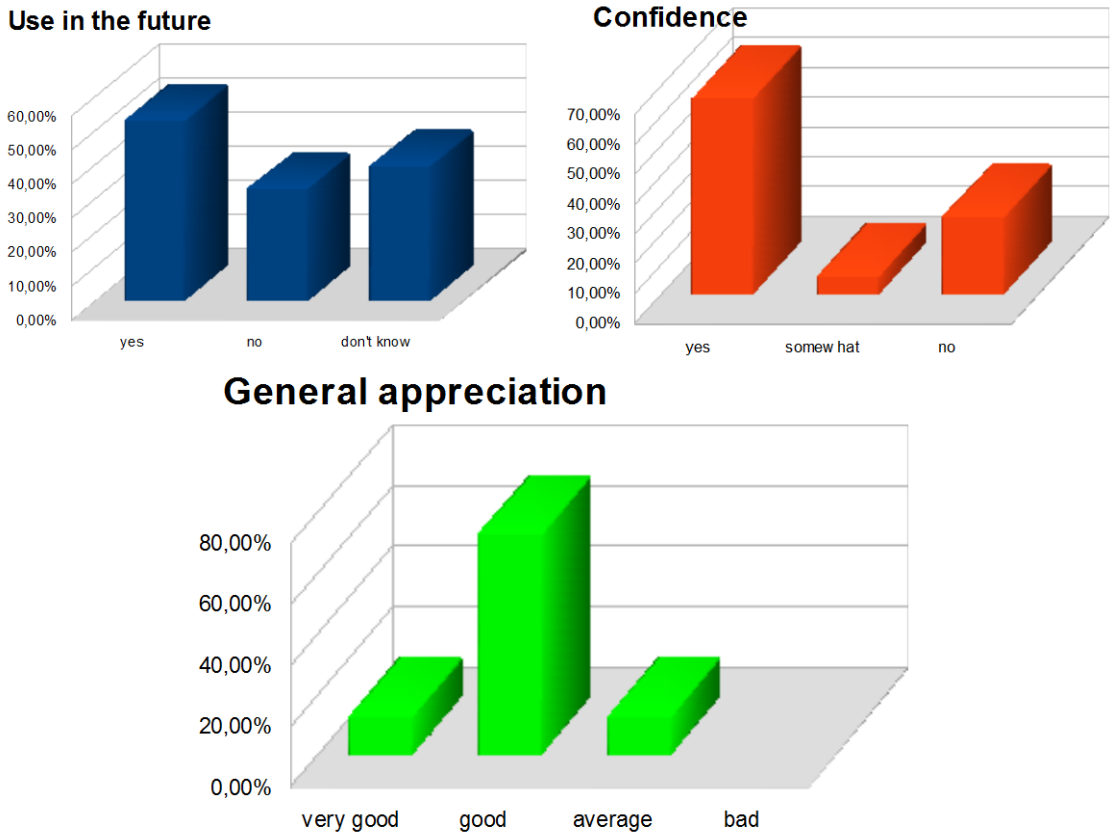
Figure 11: Evolution of the EER value given the amount of alterations.

We give now the results of the subjective evaluation experiment. We obtained for some questions similar answers for all users. As for example, all users found that the

verification is fast. 93% of them considered the system is easy to use and is non intrusive for their privacy.

Figure 12 shows the results of the subjective evaluation. Even if the biometric system is not very efficient (EER = 20.7%), nearly 50% of users are ready to use it in the future. The general appreciation is good for 60% of users.

For a logical control access, this biometric system even with a very bad value of the EER, could be an interesting solution as it necessitates none additional sensor (all computer has a keyboard), it is very simple to use. This subjective study (even if realized with a low number of individuals) shows that the perception of users is interesting to take into account.



F U T U R E T R E N D S

How can we make progress for the evaluation of behavioral biometric systems?

The constraints and the cost for the evaluation of behavioral biometric systems are extremely prohibitive. High quality benchmark databases must be available for the research community taking into account the previous constraints. These databases would facilitate the testing and the development of new behavioral biometric systems. They also would able to compare different enrollment and identification / verification algorithms to

increase the knowledge in the domain. Actually, a researcher in the domain generally creates its own database to validate the proposed system. It is generally difficult to say if the database is representative of real use cases and if the system achieves better than others in the state of the art.

The European BioSecure network of excellence (http://biosecure.it-sudparis.eu/) had for objective as for example to realize benchmark databases for different biometric modalities. If we consider behavioral modalities, only the speech was concerned. An organization should deliver for free to researchers some benchmark databases. It could be also a good thing to ask researchers to implement their biometric system following the BioAPI requirements. The cost of implementation is not so important and the benefit is high as a BSP is only a DLL file that can be transferred without giving the source code.

The statistical evaluation of biometric systems is important but is not sufficient. A biometric system to be used in real conditions must be easy to use, not reluctant to use... Subjective evaluation is a domain that needs a lot of research to take into account the user as the central element in the biometric system.

C O N C L U S I O N

We presented in this chapter several issues for the evaluation of behavioral biometric systems. We detailed the BioAPI standard that defines the architecture and the evaluation of biometric systems in a general context. Behavioral biometric modalities are currently under standardization.

Much specificity of behavioral biometric systems had been detailed in the second part of this chapter. These considerations must be taken into account for the evaluation of this kind of biometric systems by engineers or researchers in this field.

R E F E R E N C E S

Abut, H., Hansen, J.H.L., & Takeda, K., (2005). Is Our Driving Behavior Unique? *DSP for In-Vehicle and Mobile Systems*, pp. 257-274.

Adler, A., & Suckers, M. E., (2007). Comparing Human and Automatic Face recognition Performance. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 37, pp. 1248-1255.

Awad, A., & I. Traore, (2005). Detecting computer intrusions using behavioral biometrics. *3rd Annual Conference on Privacy, Security and Trust*, St. Andrews, New Brunswick, Canada, pp. 91-98.

BioApi Consortium, (2005). http://www.bioapi.org/

Cappelli, R., Maio, D., & Maltoni, D., (2002) Synthetic Fingerprint-Database Generation, *16th International Conference on Pattern Recognition (ICPR)*, Vol. 3.

Clarke, N. L., & Furnell, S. M., (2007). Advanced user authentication for mobile devices. *Computers & security*, Vol. 26, pp. 109-119.

Han, J., & Bhanu, B., (2006). Individual Recognition Using Gait Energy Image. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 28, (2), pp. 316-322.

Hanley, J. A & McNeil B. J., (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, Vol. 143, pp. 29-36.

Hemery, B., Rosenberger, C., Toinard, C., & Emile, B., (2006). Comparative study of invariant descriptors for face recognition. *8th International IEEE Conference on Signal Processing (ICSP).*

Hemery, B., Rosenberger, C., & Laurent, H., (2007) The ENSIB database: a benchmark for face recognition. *International Symposium on Signal Processing and its Applications (ISSPA), special session on "Performance Evaluation and Benchmarking of Image and Video Processing".*

Hocquet, S., Ramel, J.-Y., & Cardot, H., (2007) User Classification for Keystroke Dynamics Authentication. *International Conference on Biometrics (ICB), Lecture Notes in Computer Science 4642, Springer-Verlag Berlin Heidelberg,* pp. 531-539.

Hwang, S., Lee, H., & Cho, S., (2006). Improving Authentication Accuracy of Unfamiliar Passwords with Pauses and Cues for Keystroke Dynamics-Based Authentication. *WISI, LNCS 3917, Springer-Verlag Berlin Heidelberg,* pp. 73-78.

ISO International standard, (2006). "Information technology — Biometric performance testing and reporting" ISO/IEC 19795-1, 64 pages.

Janakiraman, R., & Sim, T., (2007) Keystroke Dynamics in a General Setting. *International Conference on Biometrics (ICB), Lecture Notes in Computer Science 4642, Springer-Verlag Berlin Heidelberg,* pp. 584–593

Mahier, J., Pasquet, M., Rosenberger, C., & Cuozzo, F., (2008). Biometric authentication, *IGI Encyclopedia of Information Science and Technology, 2nd edition*

Maltoni, M. (2004) Generation of Synthetic Fingerprint Image Databases, in N. Ratha and R. Bolle, *Automatic Fingerprint Recognition Systems*, Springer.

Muller H., Muller W., Squire D.M., Marchand-Maillet S., & Pun T., (2001). erformance evaluation in content-based image retrieval: Overview and proposals. *Pattern Recognition Letters*, Vol. 22, pp 593-601.

Muramatsu, D., & Matsumoto, T., (2007) Effectiveness of Pen Pressure, Azimuth, and Altitude Features for Online Signature Verification. *Proceedings of the International Conference on Advances in Biometrics (ICB) Lecture Notes in Computer Science 4642, Springer*, pp. 503-512.

Orozco, M., Asfaw, Y., Shirmohammadi, S., Adler, A., El Saddik, & A., (2006) Haptic-Based Biometrics: A Feasibility Study. Symposium on Haptic *Interfaces for Virtual Environment and Teleoperator Systems (HAPTICS)*, pp. 265-271.

Ortega-Garcia, J., Fierrez-Aguilar, J., Simon, D., Gonzalez, J., Faundez-Zanuy, M., Espinosa, V., Satue, A., Hernaez, I., Igarza, J.J., Vivaracho, C., Escudero D., & Moro Q.-I., (2003). MCYT baseline corpus: a bimodal biometric database, *IEEE Procedings of Image and Signal Processing*, Vol. 150, pp. 395-401.

Petrovska-Delacretaz, D., El Hannani, A., & Chollet, G., (2007) Text-Independent Speaker Verification: State of the Art and Challenges. *Lecture Notes in Computer Science, Progress in Nonlinear Speech Processing*, Vol. 4391, pp. 135-169.

Phillips, P.J., Moon, H., Rizvi, S.A., & Rauss, P.J., (2000). The FERET Evaluation Methodology for Face-Recognition Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence archive* (10), Vol. 22, pp. 1090-1104.

Sarkar, S., Phillips, P.J., Liu, Z., Vega, I.R., Grother, P., & Bowyer, K.W., (2005). The HumanID Gait Challenge Problem: Data Sets, Performance, and Analysis, *IEEE Transactions on Pattern analysis and Machine Intelligence*, Vol. 27, pp. 162-177.

Thacker, N.A. and Clark, A.F. and Barron, J.L. and Ross Beveridge, J. and Courtney, P. and Crum, W.R. and Ramesh, V. & Clark, C. (2008). Performance characterization in computer vision: A guide to best practices. *Computer Vision and Image Understanding*, Vol. 109, pp. 305-334.

T E R M S A N D D E F I N I T I O N S

Behavioral biometric: A Behavioral Biometric is a measurable behavior trait that is acquired over time for the identification or identity verification of an individual.

Benchmark: A database composed of biometric templates supposed to represent real cases for the performance evaluation of biometric systems.

Biometric Application Programming Interface (BioAPI): The BioAPI specification enables different biometric systems to be developed by the integration of modules from multiple independent companies.

Enrollment: The process of collecting biometric samples from a person and the subsequent preparation and storage of biometric reference templates representing that person's identity.

False Acceptance Rate (FAR): Rate at which an impostor is accepted by an identification system.

False Rejection Rate (FRR): Rate at which the authorized user is rejected from the system.

Equal Error Rate (EER): This error rate corresponds to the point at which the FAR and FRR cross (compromise between FAR and FRR).