

# Biometric Database Acquisition Close to “Real World” Conditions

Marcos Faundez-Zanuy<sup>1</sup>, Joan Fàbregas<sup>1</sup>, Miguel Ángel Ferrer-Ballester<sup>2</sup>,  
Aythami Morales<sup>2</sup>, Javier Ortega-García<sup>3</sup>,  
Guillermo Gonzalez de Rivera<sup>3</sup>, and Javier Garrido<sup>3</sup>

<sup>1</sup> Escola Universitària Politècnica de Mataró (Adscrita a la UPC)

<sup>2</sup> GPDS Universidad de Las Palmas de Gran Canaria

<sup>3</sup> ATVS, Universidad Autónoma de Madrid

faundez@eupmt.es

<http://www.gpds.ulpgc.es/biopassweb/biopass.htm>

**Abstract.** In this paper we present an autonomous biometric device developed in the framework of a national project. This system is able to capture speech, hand-geometry, online signature and face, and can open a door when the user is positively verified. Nevertheless the main purpose is to acquire a database without supervision (normal databases are collected in the presence of a supervisor that tells you what to do in front of the device, which is an unrealistic situation). This system will permit us to explain the main differences between what we call "real conditions" as opposed to "laboratory conditions".



## 1 Introduction

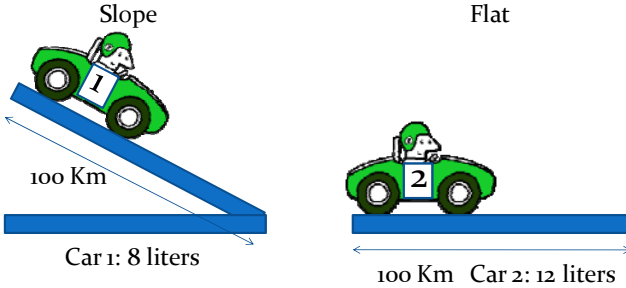
Biometric system developments are usually achieved by means of experimentation with existing biometric databases, such as the ones described in [1]. System performance is usually measured using the identification rate (percentage of users whose identity is correctly assigned) and verification errors: False Acceptance Rate (FAR, percentage of impostors permitted to enter the system), False Rejection Rate (FRR, percentage of genuine users whose access is denied) and combinations of these two basic ratios, such as Equal Error Rate (EER, or adjusting point where FAR=FRR) and Detection Cost Function (DCF) [2].

A strong problem in system comparison is that most of the times the experimental conditions of different experiments performed by different teams are not straight forward comparable. In order to illustrate this problem, let us see a simple example in the motoring sector. Imagine two cars with the fuel consumption depicted in table 1. According to this table, looking at the distance (which is equal in both cases) and the speed (which is also equal) we could conclude that car number 1 is more efficient. Nevertheless, if we look at figure 1, we realize that the experimental conditions are very different and, in fact, nothing can be concluded. This is an unfair comparison.

It is well known that car makers cannot do that. Slope, wind, etc., must be very controlled and it is not up to the car maker. Nevertheless the situation is not the same in biometrics, because there is no “standard” database to measure performance. Each fabricant can use its own database. This can let to unfair comparisons, as we explain next.

**Table 1.** Toy example for car fuel consumption comparison

|                  |   |   |
|------------------|---|---|
|                  |  |  |
| Distance         | 100 Km  | 100Km   |
| Speed            | 100 Km/h  | 100Km/h   |
| Fuel consumption | 8 liters  | 12 liters   |



**Fig. 1.** Experimental conditions corresponding to table 1

We will assume that training and testing of a given biometric system will be done using different training and testing samples, because this is the situation in real operating systems in normal life. Otherwise, this is known as “Testing on the training set”: the test scores are obtained using the training data, which is an optimal and unrealistic situation. This is a trivial problem where the system only needs to memorize the samples, and the generalization capability is not evaluated.

The comparison of different biometric systems is quite straight forward: if a given system shows higher identification rate and lower verification error than its competitor, it will be considered better. Nevertheless, there is a set of facts that must be considered, because they can let to reach a wrong conclusion.

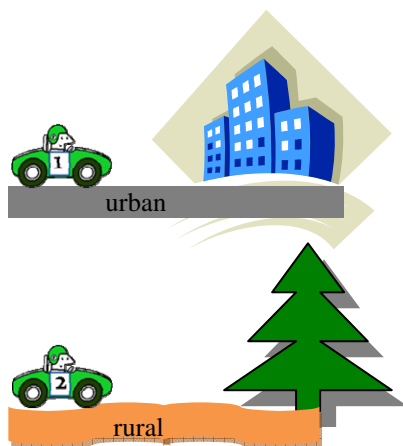
Nevertheless, there is a set of facts that must be considered, because they can let to reach a wrong conclusion. We will describe these situations in the next sections.

*A. Comparison of results obtained with different databases*

When comparing two biometric systems performing over different databases, it must be taken into account that one database can be more trivial than the other one. For instance, it does not have the same difficulty to identify people inside the ORL database [3] (it contains 40 people) than in the FERET database [4] (around 1000 people). For a study of this situation, see [5]. Thus, as a conclusion, a given system *A* performing better on Database *DB1* than another system *B* performing worse on database *DB2*, is not necessarily better, because the comparison can be unfair.

*B. Comparison of results obtained with the same database*

When comparing two biometric systems performing over the same database, and following the same protocol (same samples for training both competing systems and the remaining samples for testing), it seems that the comparison is fair. In fact it is,



**Fig. 2.** Car testing in different scenarios. The car must be the same. It does not have too much sense to design a different car for each scenario and to present experimental results in different scenarios for different cars.

but there is a problem: how can you be sure that these results will hold on when using a different database? Certainly you cannot. For this reason, researchers usually test their systems with different databases acquired by different laboratories. In the automobile example, probably, you will get the fuel consumption in several situations (urban, highway, different speeds, etc.) because one car can be more efficient in a particular scenario but it can be worse in a different one. Of course the car must be the same in all the scenarios. It will be unfair to trim the car design before making the test (one design for urban path, one design for rural path, another one for highway, etc. This would be the schematized situation in figure 2, which obviously does not have too much sense).

However, this is the usual situation in biometrics, where a new classifier is designed for each scenario. Instead of this, we propose to design a classifier, to keep it fix, and then to apply it to a different scenario (a different biometric database). The usual approaches for biometric recognition, when moving from one database to another one, imply to fit a new model because different databases contain different users. Thus, there is a risk of fine-tuning on a given database. For instance, in on-line signature recognition, [6] describes the following: “for any given database, perhaps a composite of multiple individual databases, we can always fine tune a signature verification system to provide the best overall error trade-off curve for that database –for the three databases here, I was able to bring my overall equal-error rate down to about 2.5%- but we must always ask ourselves, does this fine tune make common sense in the real world? If the fine tuning does not make common sense, it is in all likelihood exploiting a peculiarity of the database. Then, if we do plan to introduce the system into the market place, we are better off without the fine tuning.” This problematic can be illustrated in table 2.

From the example of table 2, if we only look at some results, we obtain the following conclusions:

- a) Comparing (A, DB1) with (B, DB1) we conclude that system B is better.
- b) Comparing (A, DB2) with (B, DB2) we conclude that both systems perform the same.
- c) Looking at the last column we conclude that system A is the best.

**Table 2.** Comparison of systems *A* and *B* performing on two different laboratory databases and in a real scenario

| system | DB1  | DB2 | “Real” scenario |
|--------|------|-----|-----------------|
| A      | 1%   | 1%  | 1%              |
| B      | 0.5% | 1%  | 5%              |

In which comparison is interested the system seller? Probably in the most favorable one for his/ her product. In which comparison are we (the buyers) interested? Obviously the best characterization of biometric systems is the one that we achieve with a fully operating system, where users interact with the biometric system in a “normal” and “real” way. For instance, in a door opening system, such as the system described in [7-8].

An important point is that the system must store the biometric test samples if we want to be able to repeat the experiments in the future with a different algorithm. Unfortunately, it is not easy to set up this kind of experiments and for this reason most of the research is performed in laboratory conditions. In this case, we have the risk that a well-performing system, such as system B in table 2, may be unable to generalize its good performance in a real application.

A good way to avoid the unfair comparisons represented in figures 1 and 2 is by means of international competitions such as SVC (Signature Verification Competition), FVC (Fingerprint Verification Competition), etc. In this case, all the algorithms are tested using the same evaluation protocol. Even in this case, if we fully re-train a new classifier when moving from one database to another one, we are producing a very time consuming and inefficient mechanism.

In this paper we want to emphasize the main differences between databases collected under “real conditions”, as opposed to “laboratory conditions”. This is a milestone to produce applications able to work in civilian applications. Next sections summarize the main differences between our proposed approach and classical approaches.

### 1.1 Classic Design (Step 1)

Biometric system design implies the availability of some biometric data to train the classifiers and test the results. Figure 4 on the left summarizes the flow chart of the procedure, which consists on the following steps:

1. A database is acquired in laboratory conditions. There is a human supervisor that tells the user what to do. Alternatively, in some cases, programs exist for creating synthetic databases, such as SFINGE [9] for fingerprints. Another example would be the software Faces 4.0 [10] for synthetic face generation. Figure 3 shows a synthetic fingerprint and face generated with these programs. Nevertheless, synthetic samples have a limited validity to train classifiers when applied to classify real data.

Usually databases consist of a fixed number of users with a regular number of samples per acquisition session and a concrete number of acquisition sessions. Thus, the number of available samples and the time interval between samples is quite regular and homogeneous for the whole set of users.



Fig. 3. Examples of synthetic fingerprint and face generation

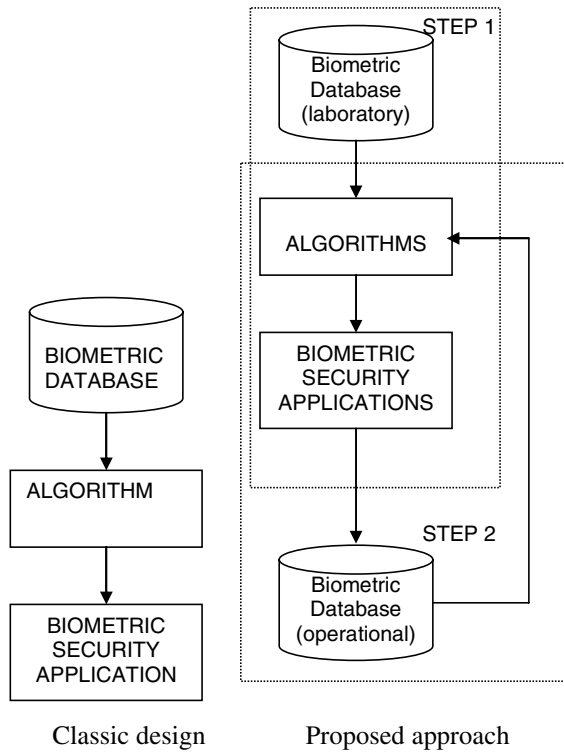


Fig. 4. Classic design (on the left) versus proposed approach (on the right)

2. After Database acquisition, a subset of the available samples is used for training a classifier, user model, etc. The algorithm is tested and trimmed using some other samples of the database (testing subset).

3. The developed system jumps from the laboratory to real world operation (physical access, web access, etc.).

This procedure is certainly useful for developing a biometric system, for comparing several different algorithms under the same training and testing conditions, etc., but it suffers a set of drawbacks, such as:

- a) In real world conditions the system will be autonomous and the user will not have chance to obtain the guidance of a human supervisor.
- b) Laboratory databases have removed those samples with low quality, because if the human supervisor detects a noisy speech recording, blurred face image, etc., will discard the sample and will ask the user for a new one. This implies that laboratory databases do not contain low quality samples. This kind of samples is useful in order to manage the failure to acquire/ failure to enroll situations.
- c) Database acquisition with a human supervisor is a time consuming task. This implies that the time interval between recording sessions and the number of samples acquired in each session tends to be quite modest.
- d) Real systems must manage a heterogeneous number of samples per user. Laboratory system developments will probably ignore this situation and thus, will provide a suboptimal performance due to mismatch between the present conditions during development and normal operation.

## 1.2 Proposed Approach (Step 2)

A more sophisticated approach involves two main steps (see figure 4 on the right). The operation can be summarized in the next steps:

1. Based on algorithms developed under the “classical approach”, a physical access control system is operated.
2. Simultaneously to system operation, biometric acquired samples are stored in a database.

This procedure provides the following characteristics:

- a) In general, the number of samples per user and the time interval between acquisitions will be different for each user. While this can be seen as a drawback in fact this is a chance to develop algorithms in conditions similar to “real world” where the user’s accesses are not necessary regular.
- b) While supervised databases contain a limited number of recording sessions, this approach permits to obtain, in an easy way, a long term evolution database.
- c) Biometric samples must be checked and labeled a posteriori, while this task is easier in supervised acquisitions.
- d) While incorrect (noisy, blurred, etc.) samples are discarded in supervised databases, they exhibit a great interest when trying to program an application able to manage the Failure to Acquire rate. In addition, these bad quality samples are obtained in a realistic situation that hardly can be obtained in laboratory conditions.

## 2 Multimodal Interface for Biometric Recognition and Database Acquisition

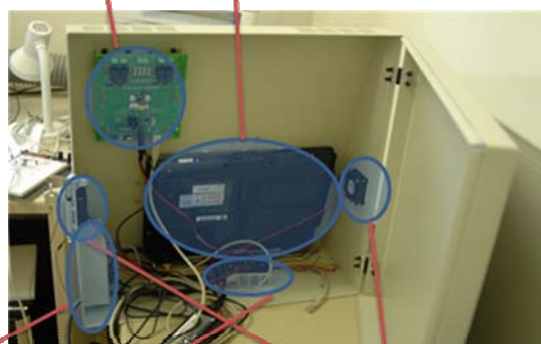
In this section we present a multimodal device specially designed to acquire speech, on-line signature, hand-geometry and face. This system has been developed under the frame of a national coordinated project between four universities, with the next main responsibilities: Universidad de Las Palmas de Gran Canaria (acquisition protocol and contact-less hand-geometry software), Universidad Autónoma de Madrid (box design, speech and on-line signature algorithms) and Escola Universitària Politècnica de Mataró (operational database collection philosophy and face recognition algorithm).

Micro:speech   Webcam:face   Webcam:hand



Touch screen: signature

IR-board   Tablet-PC



Connectors   Power supply   Fans

**Fig. 5.** Multimodal interface for biometric database acquisition (hand-geometry, speech, face and on-line signature). Frontal view (top) and rear view (bottom).

Figure 5 shows the aspect of the multimodal interface. While the system is prepared for four biometric traits, the acquisition protocol asks the user to provide his/her identity and two biometric traits (randomly selected). If both biometric traits are positively identified, the user is declared as “genuine”. In case of tilt, a third biometric trait is checked. The core of this system is a hewlett-packard notebook with touch screen (suitable for online signature acquisition). The technological solutions behind each biometric trait are DCT-NN [11] for face recognition, SVM for hand-geometry, HMM for signature and GMM for speaker recognition.

Figure 6 shows some snapshots of the screen and figure 7 shows a physical installation in a wall for door opening system.



**Fig. 6.** Some snapshots of the screen: main menu, administrative functions, and enrolment screens (on the bottom)



**Fig. 7.** Physical installation (at EUPMt) in a wall for door opening system



### 3 Real World: One Step Further from Laboratory Conditions

The goal of research should be to develop applications useful for daily usage. However, nowadays, most of the research is performed in laboratory conditions, which are far from “real world” conditions. While this laboratory conditions are interesting and necessary in the first steps, it is important to jump from laboratory to real world conditions. This implies to find a solution for a large number of problems that never appear inside the laboratory. They can be summarized in the following list:

- The user must face the system without the help of a supervisor.
- The system must be able to manage acquisitions with low quality (Failure to Acquire).
- The algorithms should consider the protocol to remove and to add users in an easy and efficient way (in a reasonable amount of time).
- The system must manage users outside the database (“open world” situation).
- The system must be able to detect coherence between training samples in order to ask for additional samples in case of troubles. In this sense, we have proposed an intelligent enrolment in [6].

In conclusion, the goal is not a fine trimming that provides a very small error in laboratory conditions. The goal is a system able to generalize (manage new samples not seen in the laboratory). It is important to emphasize that the classical Equal Error Rate (EER) for biometric system adjustment implies that the verification threshold is set up a posteriori (after knowing the whole set of test scores). While this is possible in laboratory conditions, this has no sense in a real world operation system. Thus, system performance measured by means of EER offers a limited utility.

Although a large number of different “real world” scenarios exist, and each one will presents its own particularities, most of them will have to deal with the main features described in this section.

### 4 Conclusions

In this paper we have presented a multimodal interface for biometric database acquisition. This system makes feasible the acquisition of four different biometric traits: hand-geometry, voice, on-line signature and still face image. In this paper we have emphasized the convenience of unsupervised database acquisition.

### Acknowledgements

This work has been supported by FEDER and MEC, TEC2006-13141-C03/TCM, and COST-2102.

### References

1. Faundez-Zanuy, M., Fierrez-Aguilar, J., Ortega-Garcia, J., Gonzalez-Rodriguez, J.: Multimodal biometric databases: an overview. *IEEE Aerospace and electronic systems magazine* 21(9), 29–37 (2006)

2. Martin, A., Doddington, G., Kamm, T., Ordowski, M., Przybocki, M.: The DET curve in assessment of detection performance. In: European speech Processing Conference Eurospeech 1997, vol. 4, pp. 1895–1898 (1997)
3. Samaria, F., Harter, A.: Parameterization of a stochastic model for human face identification. In: 2nd IEEE Workshop on Applications of Computer Vision, December 1994, Sarasota, Florida (1994)
4. Color FERET. Facial Image Database, Image Group, Information Access Division, ITL, National Institute of Standards and Technology (October 2003)
5. Roure-Alcobé, J., Faundez-Zanuy, M.: Face recognition with small and large size databases. In: IEEE 39th International Carnahan Conference on Security Technology ICCST 2005 Las Palmas de Gran Canaria, October 2005, pp. 153–156 (2005)
6. Jain, A.K., Bolle, R., Pankanti, S. (eds.): Biometrics, personal identification in networked society. Kluwer academic publishers, Dordrecht (1999)
7. Faundez-Zanuy, M.: Door-opening system using a low-cost fingerprint scanner and a PC. IEEE Aerospace and Electronic Systems Magazine 19(8), 23–26 (2004)
8. Faundez-Zanuy, M., Fabregas, J.: Testing report of a fingerprint-based door-opening system. IEEE Aerospace and Electronic Systems Magazine 20(6), 18–20 (2005)
9. <http://biolab.csr.unibo.it/research.asp?organize=Activities&select=&selObj=12&pathSubj=111%7C%7C12&>
10. [http://www.iqbiometrix.com/products\\_faces\\_40.html](http://www.iqbiometrix.com/products_faces_40.html)
11. Faundez-Zanuy, M., Roure-Alcobé, J., Espinosa-Duró, V., Ortega, J.A.: An efficient face verification method in a transformed domain. Pattern recognition letters 28(7), 854–858 (2007)