

Chapter 2

The BioSecure Benchmarking Methodology for Biometric Performance Evaluation

Dijana Petrovska-Delacrétaz, Aurélien Mayoue, Bernadette Dorizzi,
and Gérard Chollet

Abstract Measuring real progress achieved with new research methods and pinpointing the unsolved problems is only possible within a well defined evaluation methodology. This point is even more crucial in the field of biometrics, where development and evaluation of new biometric techniques are challenging research areas. Such an evaluation methodology is developed and put in practice in the European Network of Excellence (NoE) BioSecure. Its key elements are: open-source software, publicly available biometric databases, well defined evaluation protocols, and additional information (such as How-to documents) that allow the reproducibility of the proposed benchmarking experiments. As of this writing, such a framework is available for eight biometric modalities: iris, fingerprint, online handwritten signature, hand geometry, speech, 2D and 3D face, and talking faces.

In this chapter we first present the motivations that lead us to the proposed evaluation methodology. A brief description of the proposed evaluation tools follows. The multiple possibilities of how this evaluation methodology can be used are also described, and introduce the other chapters of this book that illustrate how the proposed benchmarking methodology can be put into practice.

2.1 Introduction

Researchers working in the field of biometrics are confronted with problems related to three key areas: sensors, algorithms and integration into fully operational systems. All these issues are equally important and have to be addressed in a pertinent manner in order to have successful biometric applications. The issues related to biometric data acquisition and the multiple issues related to building fully operational systems are not in the scope of this book, but can be found in numerous articles and books. This book is focused on the area of algorithmic developments and their evaluation. The unique feature of this book is to present and compare in one place recent algorithmic developments for main biometric modalities with the proposed evaluation methodology.

Biometrics could be seen as an example of the pattern recognition field. Biometric algorithms are designed to work on biometric data and that point introduces a series of problems related to biometric databases. More data is always better from the point of view of pattern recognition. Databases are needed in the different phases of the design of classifiers. More development data usually leads to better classifiers. More evaluation data (leading to a bigger number of tests) gives statistically more confident results. Databases should be also acquired in relation to the foreseen application, so for each new system or application, new data have to be acquired.

But collecting biometric samples is not a straightforward task, and in addition to a number of practical considerations, it also includes issues related to personal data protection. The personal data protection laws are also different between countries, as explained in [14]. All these issues result in the fact that the collection and availability of relevant (multi site and multi country) biometric databases is not an easy task, as explained in [6].

Once the problem of availability and pertinence of the biometric databases is solved, other problems appear. Biometrics technologies require multidisciplinary interactions. Therefore collaborative work is required in order to avoid duplication of efforts. One solution to avoid duplication of efforts in developing biometric algorithms is the availability of open-source software. That is one of the concerns of our proposal for a research methodology based on the usage of open-source software, and publicly available databases and protocols, which will enable fair comparison of the profusion of different solutions proposed in the literature. Such a framework will also facilitate reproducibility of some of the published results, which is also a delicate question to address. These reproducible (also denoted here as benchmarking) experiments can be used as comparison points in different manners.

In this chapter we will introduce in more detail the proposed benchmarking framework using some hypothetical examples to illustrate our methodology. The benchmarking framework is designed for performance evaluation of monomodal biometric systems. It can also be used for multimodal experiments. In the following chapters, the proposed framework is put in practice for the following biometric modalities: iris, fingerprint, hand geometry, online handwritten signature, speech, 2D and 3D face, talking faces, and multimodality. The algorithms (also called systems) that are analyzed within the proposed evaluation methodology are developed in the majority of the cases by the partners of the European Network of Excellence (NoE) BioSecure [3]. The major part of the work related to this framework was realized in the framework of the BioSecure NoE supported to better integrate the existing efforts in biometrics.

2.2 Terminology

In order to explain our ideas we will first introduce some definitions that will be used through this book. Figure 2.1 is an illustration of a generic biometric experiment. Researchers first develop their biometric algorithms. In the majority of cases

they need development data. A typical example for the need of development data is the need for face images to construct the face space with the Principal Component Analysis (PCA) method. Data are also needed to fix thresholds, to tune parameters, to build statistical models, or to learn fusion rules. Ideally, researchers should aim at developing systems that could have generalization capabilities. In this case these development data should be different from the data used for evaluation. But for new emerging biometrics it is not easy to have at disposal enough data for disjoint development and evaluation partitions.

The biometric data necessary for the development part are denoted here as *Development data*, or *Devdb*. The *biometric algorithms*, or *pattern recognition algorithms* for sake of simplicity are also denoted as *systems*, and have not to be taken in their more generic usage as a system designing the whole biometric application. The performance of these *systems* is measured (evaluated) on *Evaluation database and protocols* (abbreviated as *Evaldb*) and give a certain result, denoted as *Result System X*.

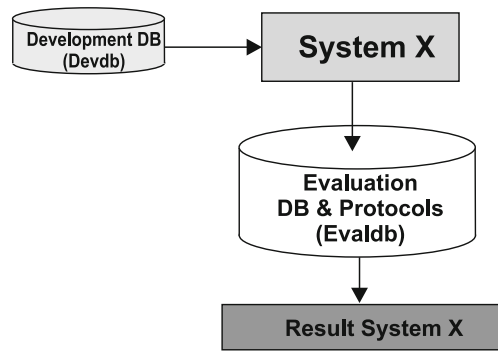


Fig. 2.1 Flow diagram of a generic biometric experiment. *Devdb* stands for Development database, *Evaldb* for Evaluation database and protocols. When *System X* uses this data for the development and evaluation part, it will deliver a result, denoted as *Result System X*

2.2.1 When Experimental Results Cannot be Compared

The majority of published work where researchers report their biometric experiments could not be compared to other published work in the domain. There are multiple reasons of this noncomparability of published work, such as:

- The biometric databases that underlie the experiments are private databases.
- The reported experiments are done using publicly available databases, but evaluation protocols are not defined, and everybody can choose his own protocol.
- The experiments use publicly available databases (with available evaluation protocols), but in order to show some particularities of the proposed system, the researchers use “only” their evaluation protocol, adapted to their task.

This situation is illustrated in Fig. 2.2 where the reported results of the three biometric systems are not comparable. This situation is found in the majority of current scientific publications regarding biometrics. Let us assume that there are three

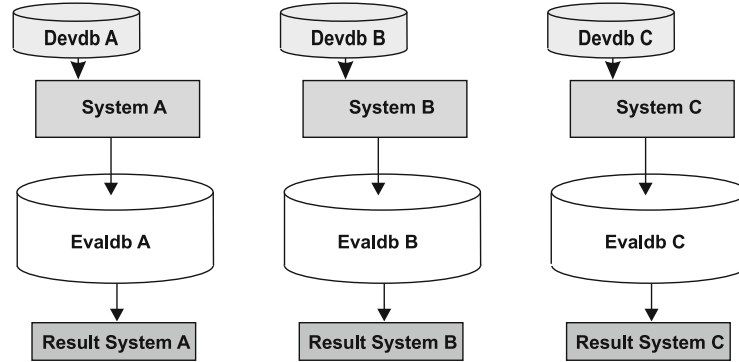


Fig. 2.2 Current research evaluation status with no common comparison point. *Devdb* stands for Development database and *Evaldb* for Evaluation database and protocols

research groups coming from three different institutions reporting results on their new algorithms that they claim provide the solution to the well known illumination problem for face verification. *Group A* has a long experience in the domain of image processing and indexing, and has on its hard disks a lot of private and publicly available databases. *Group B* has a lot of experience, but not as many databases, and *Group C* is just starting in the domain. All of them are reporting results in the same well known international conference and all of them claim that they have achieved significant improvements over a baseline experiment. Let us further assume that the baseline experiment that the three groups are using is their own implementation of the well known Principal Component Analysis (PCA) algorithm, with some specific normalizations of the input images. Let us further guess what the main characteristics of their systems are:

- *Group A* is comparing their new research results with a statistical method using Hidden Markov Models–HMMs (denoted as *System A*) in Fig. 2.2. They report results they obtained on an evaluation database and protocol (abbreviated here as *EvalDb A*). They report that the database (publicly available) is composed of 500 subjects, with different sessions with a maximum temporal variability of one year between the sessions, and with a certain kind of illumination variability (the illumination source is situated on the extreme left side of the face). They report that they are using some private databases (denoted as *Devdb A*) for their development part, more precisely to build their HMM representing their “world model.” They report that no overlap of subjects is present between development and evaluation databases. They report 30% relative improvement over the baseline PCA based algorithm with their research *System A*. No precise details on the normalization procedure of the face images are given, aside from the fact that as a first step they use histogram equalization.
- *Group B* is trying to solve the problem in a different way. They are working on image enhancement methods. They use two publicly available development databases, and report results on their private database. Their database is not

publicly available, has 50 subjects, no temporal variability, and a huge variability of illumination. They report 40% of relative improvement over their baseline PCA-based implementation with their research *System B*. They do not give details if they use histogram equalization prior to their image enhancement method. The rest of their algorithms is the same as their baseline PCA system.

- *Group C* is claiming to have 60% relative improvement with their new method, that we call here *System C*. Their baseline is a proprietary PCA implementation, and their new research algorithm is based on the Support Vector Machine (SVM) method for the classification (matching) step. They use sub sets of two publicly available databases for development and evaluation with no details regarding the experimental protocols.

What conclusions could we make if we try to summarize their results and use their findings in order to understand which method is the best, or which part of their algorithms is responsible for the improvements that they report? Without being able to reproduce either of the experiments, the conclusion is that each of the above cited groups has made great improvements, but that their results cannot be compared among them. They have certainly developed better solutions tailored for their evaluation databases, but these experiments do not guarantee that the results could be used and/or generalized to another databases. A lot of questions arise:

- What if the PCA baseline method they have used is a quickly developed version which has a bad performance?
- What if one could also achieve improvements using a better tuning of the PCA baseline method?
- What if by using well-tuned baseline PCA algorithms with a simple histogram equalization in the pre-processing step we get almost as good of results as we do with the more complicated and computationally more expensive methods?
- What about the classification power of the SVM?
- What about the combination of the feature extraction step of *Group B* with the classification method of *Group C*?

Finally, the conclusion is that *Results A*, *B*, and *C* cannot be compared. We cannot conclude which is the best method or what the major contributions of each method are.

2.2.2 Reporting Results on a Common Evaluation Database and Protocol(s)

Hopefully, there are some common evaluations proposed and used in the field of biometrics. The basic scheme of such evaluation campaigns is depicted in Fig. 2.3. Let us suppose that the evaluation database and protocols are given to the participants, on the condition that they submit their results, and publish only their own results.

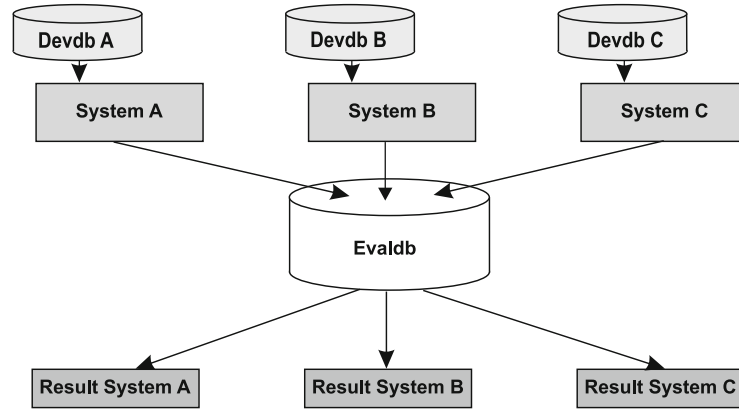


Fig. 2.3 Current research evaluation status with common evaluation database and protocols

We will continue to follow our hypothetical *Groups A, B, and C*. Let us further assume that they have submitted slightly modified versions of their *Systems A, B, and C* in another well-known conference (six months later), describing their submitted versions on the Evaluation Campaign, and each one in a separate paper. *Group A* is first, *Group B* holds the third position and *Group C* is in seventh place, among 10 participating institutions. But this ranking is only available to the participants of the campaign, and each institution can publish only their own results. Compared to their previous publications, this time they have used a common evaluation database and protocol, keeping secret their development databases. The evaluation database is a new database made available only to the participants of the evaluation campaign.

The researchers who have not participated in the evaluation campaign do not know what the best results obtained during this evaluation campaign are. What conclusions can be made when comparing their three biometric systems, through the three published papers? That using HMM modelling is best suited for solving the problem of illumination variability posed by this evaluation? Could these results be reproduced by another laboratory that does not have all the development databases that are owned by *Group A*? Or is it the normalization module of *Group B* that is mostly contributing to their good results? Or could it be that the image enhancement technique from *Group B*, when combined with the classification method of *Group C*, would give even better results? Or could it be that it is the HMM modelling method in combination with the image enhancement techniques that would give the best results?

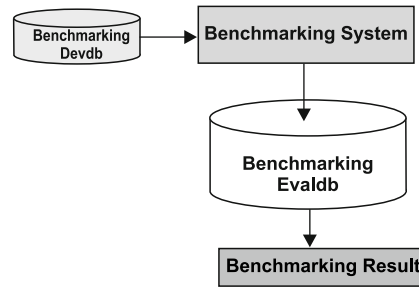
The conclusion is that for that particular campaign, and only for institutions that have participated, they can say which institution has the best results. And their results are still dependent on the database and protocol.

There is also an ambiguity with the previously reported results of the same institutions. Are the systems described by *Group A* in different publications the same? In this case the difference in performance should be only related to the database. Or is it that they have modified some parts? And if yes do we know exactly which ones? Are the results reproducible?

2.2.3 Reporting Results with a Benchmarking Framework

Let us assume that some organization has prepared an evaluation package including a publicly available (at distribution cost) database, mandatory and auxiliary evaluation protocols, and a baseline open-source software. Let us also assume that the database is divided in two well-separated development and evaluation sets. We will denote this package as a *Benchmarking package*. The building components of this Benchmarking (or Reference) Framework (or Package) are schematically represented in Fig. 2.4.

Fig. 2.4 Building components of a benchmarking framework (package)



If the default parameters of the software are well chosen and use the predefined development data, the results of this baseline system can be easily reproduced. In order to avoid using the baseline software with parameters that are not well chosen, leading to bad results, a How-to document could be of great help. Such benchmarking experiments could serve as a baseline (comparison point), and measure the improvements of new research systems.

The above cited components compose the benchmarking (also called reference) framework. Such a benchmarking or reference framework should be composed of

- Open-source software.
- Publicly available databases.
- Disjoint development and evaluation sets, denoted as *Devdb Bench* and *Evaldb Bench*.
- Mandatory evaluation protocols.
- How-to documents that could facilitate reproducibility of the benchmarking experiment(s).

The algorithms implemented in this benchmarking framework could be baseline or state-of-the-art algorithms. The distinction is rather subtle and is related to the maturity of the research domain. For mature technologies normally a method is found that has proven its efficiency and best performance. As such state-of-the-art algorithms we can mention the minutia based fingerprint systems, or the usage of Gaussian Mixture Models (GMM) for speaker verification. It should be noted that a distinction should be made between a method and its implementation. If the

convergence point is not reached, then the systems are denoted as baseline, related to a well known algorithmic method (for example the PCA for face analysis). In the rest of our document the distinction is sometimes made between baseline and state-of-the-art algorithms, but in the majority of the cases the software implementations are denoted as comparison, reference or benchmarking software (algorithms or system).

Their purpose is to serve as a comparison point and should be based on principles that have proven their efficiency. They should be modular and be composed of the major steps needed in a biometric system, such as preprocessing, feature extraction, creation of models and/or comparison and decision making.

Our major concern when proposing, describing, illustrating and making available such a framework for the major biometric modalities is to make a link between different published work. This link is the benchmarking framework that is introduced and put in practice in this book. The main component of this evaluation methodology is the availability and definition of benchmarking experiments, depicted in Fig. 2.4. This benchmarking package should allow making a link between different publications in different ways.

It has to be noted that it is the methodology that we are mainly concerned with and not the software implementation. Therefore “framework” does not design an informatics environment with plug and play modules, but rather autonomous C or C++ software modules with well-defined input and output data. Results of modules could be easily combined with other software. This is necessary when dealing with research systems that have the potential to change very fast, and are not here to serve as a practical implementation of a well proven method.

Let us continue to follow our three groups that decided to use such a framework. Young researchers arriving in these groups could immediately be faced with state-of-the-art or debugged baseline systems and they could put all their efforts into implementing their new ideas, rather than trying to tune the parameters in order to obtain satisfactory results of the baseline systems. We will follow the new PhD student that just arrived in *Group A*, who would like to fully concentrate his work on a new feature extraction method in order to improve current feature extraction parameters used for speaker verification. He would like to investigate in the direction of replacing the spectral parameters that are widely used for speech and speaker recognition by using wavelets. Using a benchmarking framework should avoid any loss of time in developing all the components of a state-of-the-art system from scratch. For mature technologies developing such systems requires a lot of work, personnel, and experience, and could not be done in a short time.

Another new PhD student arrived in *Group B*, and it happens that he would like to work on the same problem as PhD A, but with another features. Using the same benchmarking framework would allow them to work together and to constructively share their newly acquired findings and to measure their results against a mature system.

Such a benchmarking framework is introduced in this book, for the major biometric modalities. It is not only described, but the open-source programs have been developed, tested and controlled by persons other than the developers. The

benchmarking databases and protocols have been defined and results of the benchmarking framework (benchmarking results) made available, so that they could be fully reproduced. A short description of the benchmarking framework is given in the following section.

2.3 Description of the Proposed Evaluation Framework

The proposed framework was developed by the partners of the BioSecure NoE. The evaluation framework concerns eight biometric modalities. It is comprised of publicly available databases and benchmarking open-source reference systems (see Table 2.1). Among these systems, eight were developed within the framework of the NoE BioSecure. We also included the existing fingerprint and speech open-source software. The added value for those existing open-source software is the fact that we defined benchmarking databases, protocols and How-to documents, so that the benchmarking experiments could be easily reproduced.

Table 2.1 Summary of the BioSecure Evaluation Framework [8]. *ICP* stands for Iterative Closest Point, and *TPS* stands for Thin Plate Spline warping algorithm

Modality	System	Short description	Database(s)
Iris	BioSecure	Inspired by Daugman's algorithms [5]	CBS [8]
Fingerprint	NFIS2 [11]	Minutiae based approach	MCYT-100 [13]
Hand	BioSecure	Geometry of fingers	BioSecure [8], BIOMET [8]
Signature	BioSecure Ref1	Hidden Markov Model (HMM)	MCYT-100 [13], BIOMET [8]
Signature	BioSecure Ref2	Levenshtein distance measures	MCYT-100 [13], BIOMET [8]
Speech	ALIZE [1]	Gaussian Mixture Models (GMM)	BANCA [2], NIST'2005 [12]
Speech	BECARS [10]	Gaussian Mixture Models (GMM)	BANCA [2]
2D face	BioSecure	Eigenface approach	BANCA [2]
3D face	BioSecure	ICP and TPS warping algorithms	3D-RMA [4]
Talking-face	BioSecure	Fusion of speech and face software	BANCA [2]

The material related to this framework could be found on the companion URL [8]. This framework is also used for the experiments reported in the rest of the chapters of this book. Each chapter includes a section related to the benchmarking framework of the modality in question. In those sections more detailed description of the methods implemented in the benchmarking software are given, as well as reasons about the choice of the benchmarking database (with partitions including disjoint development and evaluation sets), including descriptions of the mandatory protocols that should be used as well as the expected results.

In the following paragraph, the collection of existing (at the time of writing) material [8] is enumerated.

1. **The Iris BioSecure Benchmarking Framework:** *the Open Source for IRIS (OSIRIS) v1.0 software was developed by TELECOM SudParis. This system is*

deeply inspired by Daugman algorithms [5]. The associated database is the CBS database [8]. The system is classically composed of a segmentation and classification steps. The segmentation part uses the canny edge detector and the circular Hough transform to detect both iris and pupil. The classification part is based on Gabor phase demodulation and Hamming distance measure.

2. **The Fingerprint BioSecure Benchmarking Framework:** *the open-source software is the one proposed by NIST [11], NFIS2-rel.28-2.2. The database is the bimodal MCYT-100 database [13], with two protocols (one for each sensor data).* It uses a standard minutiae approach. The minutiae detection algorithm relies on binarization of each grayscale input image in order to locate all minutiae points (ridge ending and bifurcation). The matching algorithm computes a match score between the minutiae pairs from any two fingerprints using the location and orientation of two minutiae points. The matching algorithm is rotation and translation invariant.
3. **The Hand BioSecure Benchmarking Framework:** *the open-source software v1.0 was developed by Epita and Bo  azi   University and uses geometry of fingers for recognition purposes [7]. The hand database is composed of three databases: BioSecure, BU (Bo  azi   University) and the hand part of the BIOMET [8], with protocols for identification and verification experiments.* The hand image is first binarized. The system searches the boundary points of the whole hand and then the finger valley points. Next, for each finger its major axis is determined and the histogram of the Euclidian distances of boundary points to this axis are computed. The five histograms are normalized and are thus equivalent to probability density functions. These five densities constitute the features used for the recognition step. Thus, given a test hand and an enrollment hand, the symmetric Kullback-Leibler distance between the two probability densities is computed separately for each finger. Only the three lower distance scores are considered and summed yielding a global matching distance between the two hands.
4. **The Online Handwritten Signature Benchmarking Framework:** *the signature part is composed of two open-source software Ref1 and Ref2, two databases (the signature parts of MCYT-100 [13] and BIOMET [8] databases, accompanied by six mandatory protocols).*
 - a. *The first signature open-source software, denoted here as Ref1 v1.0 (or Ref-HMM) was developed by TELECOM SudParis. It uses a continuous left-to-right Hidden Markov Model (HMM) to model each signer's characteristics [17]. 25 dynamic features are first extracted at each point of the signature. At the model creation step, the feature vectors extracted from several client signatures are used to build the HMM. Next, two complementary information derived from a writer's HMM (likelihood and Viterbi score) are fused to produce the matching score.*
 - b. *The second signature open-source software, denoted here as Ref2 v1.0 or Ref-Levenshtein, was developed by University of Magdeburg and it is based on Levenshtein distance measures [15]. First, an event-string modelling of features derived from pen-position and pressure signals is used to represent*

each signature. In order to achieve such a string-like representation of each online signature, the sample signal is analyzed in order to extract the feature events (such as gap between two segments, local extrema of the pressure, pen-position in x – axis and y – axis, and velocity) which are coded with single characters and arranged in temporal order of their occurrences leading to an event string (which is simply a sequence of characters). Then, the Levenshtein distance is used to evaluate the similarity between a test and a reference event string.

5. **The Speech Evaluation Framework:** *is composed of two open-source software ALIZE v1.04 [1] and BECARs v1.1.9 [10]. Both ALIZE and BECARs software are based on the Gaussian Mixture Model (GMM) approach. The publicly available BANCA [2] database (the speech part), and the NIST'05 database [12] are used for the benchmarking experiments.* The speech processing is done using classical cepstral coefficients. A bi-Gaussian modelling of the energy of the speech data is used to discard frames with low energy (i.e., corresponding to silence). Then, the feature vectors are normalized to fit a zero-mean and a unit variance distribution. The GMM approach is used to build the client and world models. The score calculation is based on the estimation of the log-likelihood between the client and world models.
6. **2D Face Benchmarking Framework:** *the open-source software v1.0 was developed by Boğaziçi University and it uses the standard eigenface approach [16] to represent faces in a lower dimensional subspace. The associated database are the extracted 2D images from BANCA [2]. The mandatory protocols are the P and Mc protocols.* The PCA algorithm works on normalized images (with the detected positions of the eyes). The face space is built using a separate training set (from the Devdb) and the dimensionality of the reduced space is selected such that 99% of the variance is explained by the Principal Component Analysis (PCA). At the feature extraction step, all the enrollment and test images are projected onto the face space. Then, the L_1 norm is used to measure the distance between the projected vectors of the test and enrollment images.
7. **3D Face BioSecure Benchmarking Framework:** *the open-source software v1.0 was developed by Boğaziçi University [9]. The 3D-RMA database is used for the benchmarking experiments.* The 3D approach is based on the Point Set Distance (PSD) technique and uses both Iterative Closest Point (ICP) and Thin Plate Spline (TPS) warping algorithms. A mean 3D face has to be constructed as a preliminary step. Then facial landmark coordinates are automatically detected on each face (using the ICP algorithm). Then using the TPS algorithm the landmarked points of each face move to the coordinates of their corresponding landmarks of the mean face. After that, the warped face is re-sampled so that each face contains equal number of 3D points. Finally, similarities between two faces are calculated by Euclidean norm between registered 3D point sets.
8. **The Talking Face BioSecure Benchmarking Framework:** *is a fusion algorithm for face and speech. The 2D Face BioSecure software is used for the face part, and for the speech part the same two open-source software as for the*

speech modality are chosen. The database is the audio-video part of the BANCA database [2]. The min-max approach is used to fuse the face and speech scores.

2.4 Use of the Benchmarking Packages

The evaluation packages can be found on a companion URL. A public access (for reading/downloading only) is available via a Web interface at this URL

`http://share.int-evry.fr/svnview-eph/`

The following material is available for the eight biometric modalities reported in the previous section. The source code of the reference systems described in Sect. 2.3, scripts, Read-mes, How-tos, lists of tests (impostor and genuine trials) to be done, and all the necessary information needed to fully reproduce the benchmarking results are available at this address. In the How-to documents more details about the following points are to be found:

- Short description of the reference database (and a link to this publicly available database).
- Description of the reference protocol.
- Explanation about the installation and the use of the reference system.
- The benchmarking results/performance (including statistical significance) of the reference system using the reference database in accordance with the reference protocol.

On the companion URL, additional open-source software can be found, which could be useful for running biometric experiments. This additional software is used, for example, for eye detection (when face verification algorithms require the position of the eyes), or they explain how to calculate the confidence intervals, or how to plot the results curves.

In practice, the user of our evaluation tool can be confronted with the following scenarios:

1. He wants to reproduce the benchmarking results to be convinced of the good installation and use of the reference system, and also do some additional experiments by changing some parameters of the system, in order to become familiar with this modality. He should
 - Download the reference system and the associated database.
 - Compile the source code and install the reference system.
 - Run the reference system on the reference database using the scripts and list of trials provided within the evaluation framework.
 - Verify that the obtained results are the same as the benchmarking results.
 - Run some additional test.

2. He wants to test the reference system on his own database according to his own protocol in order to calibrate this particular task. In this case, he has to process as follows:
 - Download the reference system from the SVN server.
 - Compile the source code and install the reference system.
 - Run the reference system using his own database and protocol (for this task, he is helped by the script files provided within the evaluation framework).
 - Compare the obtained results to those provided within the evaluation framework to evaluate the “difficulty” of these particular database and protocol.
3. He wants to evaluate his new biometric algorithm using the reference database and protocol to calibrate the performance of his algorithm. In this case, he has to process as follows:
 - Download the reference database.
 - Run the evaluated system on the reference database using the list of trials provided within the evaluation framework.
 - Compare the obtained results to the benchmarking results provided within the evaluation framework to evaluate the performance of the new algorithm.

These scenarios can also be combined. Running more experiments is time consuming, but the results of comparing a newly developed system within different evaluation scenarios is more valuable. In such a way the research system can not only be compared to the reference system, but also on other databases and with new experimental protocols. All these comparisons will give useful information about the behavior of this system in different situations, its robustness, how competitive it is compared to the reference system and to the other research systems that have used the same evaluation methodology.

2.5 Conclusions

This framework has been widely used in this book to enable comparisons of the algorithmic performances presented in each of the next nine chapters.

References

1. ALIZE: a free and open tool for speaker recognition. <http://www.lia.univ-avignon.fr/heberges/ALIZE/>.
2. E. Bailly-Baillière, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Mariéthoz, J. Matas, K. Messer, V. Popovici, F. Porée, B. Ruiz, and J.-P. Thiran. The BANCA Database and Evaluation Protocol. In *4th International Conference on Audio-and Video-Based Biometric Person Authentication (AVBPA'03)*, volume 2688 of *Lecture Notes in Computer Science*, pages 625–638, Guildford, UK, January 2003. Springer.

3. BioSecure Network of Excellence. <http://biosecure.info/>.
4. 3D RMA database. http://www.sic.rma.ac.be/~beumier/DB/3d_rma.html.
5. J. G. Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE Trans. Patt. Ana. Mach. Intell.*, 15(11):1148–1161, 1993.
6. P. J. Flynn. Biometric databases. In A. Jain, P. Flynn, and A. Ross, editors, *Handbook of Biometrics*, pages 529–548. Springer, 2008.
7. G. Fouquier, L. Likforman, J. Darbon, and B. Sankur. The Biosecure Geometry-based System for Hand Modality. In *the Proceedings of 32nd International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Honolulu, Hawaii, USA, april 2007.
8. BioSecure Benchmarking Framework. <http://share.int-evry.fr/svnview-eph/>.
9. M. O İrfano  lu, B. G  kberk, and L. Akarun. Shape-based Face Recognition Using Automatically Registered Facial Surfaces. In *Proc. 17th International Conference on Pattern Recognition*, Cambridge, UK, 2004.
10. BECARS Library and Tools for Speaker Verification. <http://www.tsi.enst.fr/becars/index.php>.
11. National Institute of Standards and Technology (NIST). <http://www.itl.nist.gov/iad/894.03/fing/fing.html>.
12. NIST Speaker Recognition Evaluations. <http://www.nist.gov/speech/tests/spk>.
13. J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, M. F. J. Gonzalez, V. Espinosa, A. Satue, I. Hernaez, J. J. Igarza, C. Vivaracho, D. Escudero, and Q. I. Moro. MCYT baseline corpus: A bimodal biometric database. *IEE Proceedings Vision, Image and Signal Processing, Special Issue on Biometrics on the Internet*, 150(6):395–401, December 2003.
14. M. Rejman-Greene. Privacy issues in the application of biometrics: a european perspective. volume Chapter 12. Springer, 2005.
15. S. Schimke, C. Vielhauer, and J. Dittmann. Using Adapted Levenshtein Distance for On-Line Signature Authentication. In *Proceedings of the ICPR 2004, IEEE 17th International Conference on Pattern Recognition, ISBN 0-7695-2128-2*, 2004.
16. M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.
17. B. Ly Van, S. Garcia-Salicetti, and B. Dorizzi. On using the Viterbi Path along with HMM Likelihood Information for On-line Signature Verification. *IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics, Special Issue on Recent Advances in Biometric Systems*, 37(5):1237–1247, October 2007.

<http://www.springer.com/978-1-84800-291-3>

Guide to Biometric Reference Systems and
Performance Evaluation

Petrovska-Delacrétaz, D.; Chollet, G.; Dorizzi, B. (Eds.)

2009, XXX, 382 p. 146 illus., 15 illus. in color.,

Hardcover

ISBN: 978-1-84800-291-3