Chapter 2
Data Representation and Analysis

2.1 Introduction

The last few years have witnessed the emergence of new tools and means for the scientific analysis of image-based information for security and forensic science and crime prevention applications. For instance, images can now be captured, viewed and analysed at the scenes or in laboratories within minutes whilst simultaneously making the images available to other experts via fast and secure communication links on the Internet, thereby making it possible to share information for forensic and security intelligence and crime linking purposes. In addition, these tools have a strong link with other aspects of investigation, such as image capture, information interpretation and evidence gathering. They are helpful for minimization of human error and analysis of data. Although there exist a number of application scenarios, the analysis of data is usually based on a conventional biometric system. Therefore, the following discussion on a biometric system is given as it would be a starting point for any other imaging system for use in security and/or forensic science.

A standard Biometric Identification System consists of the following three phases: Data Acquisition, Feature Extraction and Matching, and operates in two distinct modes: Enrolment Mode or Identification Mode [1]. The Data Acquisition stage is used in the enrolment mode to establish the database of users and their related biometric data whereas in Identification mode it is used to obtain a reference biometric from the user. This reference biometric is then processed at the Feature Extraction phase to obtain unique and comparable features. These features are then compared in the Matching phase with the related features of all the biometric templates in the database to establish or refute the identity of the user. Figure 2.1 depicts a block diagram view of a basic Biometric Identification System.

The design of any biometric system is based on decisions regarding the selection of appropriate modules for each of these processes [1, 3]. Details of these processes and modules included within these processes along with the critical issues that need to be addressed before a design decision is made are described below.
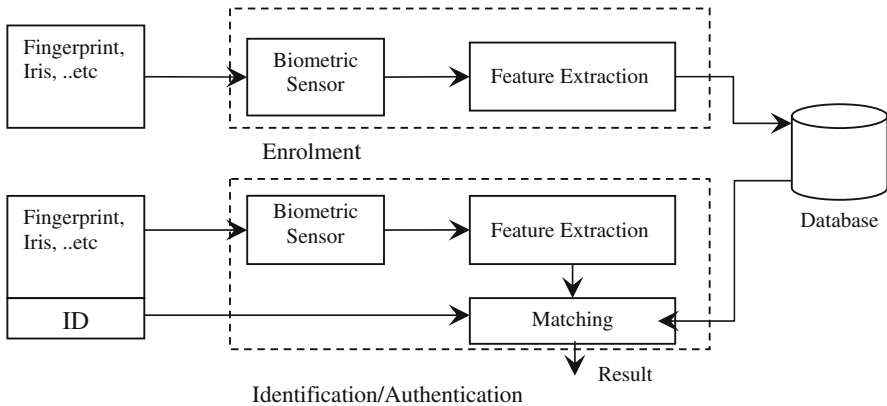
Fig. 2.1 Block diagram of a biometric identification system

2.2 Data Acquisition

In the data acquisition process the first and foremost decision to make pertains to the selection of an appropriate biometric trait. A lot of thought has to go into the selection of human physical and physiological traits for use by the biometric recognition system. The selected trait has to be universal but unique so that the trait exists in all users but also varies from subject to subject. It has to be permanent and resistant to changes so that the stored biometric data is usable over a long period of time. It has to be measurable and socially and economically acceptable so that the data can be gathered and matching can be performed and results quantified within a reasonable time and cost constraints. It should also be very difficult to circumvent or forge the trait. In addition, attention should be paid to ensure that machine-readable representations should completely capture the invariant and discriminatory information in the input measurements. This representation issue is by far an important requirement and has far reaching implications on the design of the rest of the system. The unprocessed measurement values should not be invariant over the time of capture and there is a need to determine those peculiar/salient features of the input measurement which both discriminate between the identities as well as remain invariant for a given individual.

The acquisition module should also aim to capture salient features since it is accepted that more distinctive biometric signals offer more reliable identity authentication while less complex measurement signals inherently offer a less reliable identification results. Therefore, quantification at an earlier stage would lead to much improved and effective results of a biometric recognition system.

It is important to note that no single biometric is expected to meet all the above mentioned requirements. Therefore, developers are required to identify the best possible biometric trait for each application e.g. for an application that

Table 2.1   Some commonly used biometrics

Physical	Behavioural
Fingerprint:	**Gait:**
Most commonly used	Useful in distant Surveillance
Higher false accept rate	Changes with age and surface
Face:	**Voice:**
Easiest to acquire	Useful in absence of visual data
Difficult to compar	Changes with age and Health
Hand Geometry:	**Handwriting & Signature:**
Robust under different conditions	Useful in detecting emotions as well
Changes occur with age	Changes with age, health and stress
Palm Print:	
Bigger area of interest	
Availability of data set	
Iris:	
Very Low false accept rate	
Difficult to acquire	
Ear:	
Robust to change	
Difficult to acquire	

focuses on access control to critical area or application cost may not necessitate a significant consideration but uniqueness and circumvention may be important. Some of the most commonly used biometric traits are identified in the following Table 2.1.

2.2.1 Sensor Module

Once a biometric trait is identified a suitable sensor module is required to capture the data. The sensor module is the interface between the user and the system, therefore the selection of an appropriate sensor module is critical for the success of the system. A poor quality sensor module would not be able to capture the biometric data properly, thus increasing the failure to capture error. This will cause user dissatisfaction and low acceptance rate which may eventually cause the system to fail.

It is also important to decide if the sensor module would be "overt" or "covert". If the user is aware of the fact that biometric identification is taking place the system is called overt but if they are not aware that any identification is taking place then the system is called covert. Overt systems are mostly used in access control applications whereas covert systems are used in surveillance and continuous monitoring systems. For example, a computer may contain a fingerprint scanner to identify the user at logon (which is the example of an overt system) but then constantly keep acquiring facial images from an attached webcam to verify that the same user that logged on is still accessing the system (which is the example of a covert system).

As most biometric systems are imaging based, the quality and maintenance of the raw captured biometric image also plays an important role in the development of a strong biometric identification system.

2.2.2 Data Storage

Maintenance of the template data store 'or' template database is often an over looked area of the biometric identification system. A well established, secure and effective database can improve the performance and user acceptance of the system. As the database has to store the biometric data along with other personal details of the user, it has to be kept very secure. The size of the database also has to be kept as small as possible to maintain the speed of access.

The type of data to be stored depends upon the kind of application that will utilize the data. Raw images are stored for research and feature sets are usually stored for real world applications. Both type of data storage provides some interesting challenges [4].

2.2.2.1 Raw Images

Raw Images are stored for research and development applications. When storing raw images the following points have to be kept in mind.

Image Size – Image Size should be kept small to reduce the database size but if the image size is very small a lot of important information may be lost.

Image Format – As the data is being stored for research it is important to keep the images in a standard file format so that they can also be accessed by other researchers. The image should be stored in a format that does not use a strong compression algorithm because compression can cause change to pixel values in the image. This can cause corruption of data.

Image DPI – A high DPI image will have more information available and thus will provide more data for researchers but on the other hand it will also have a large size thus increasing the size of the database. A balance needs to be obtained between the DPI and Image size for optimal performance of the database.

Usually an image with a size of around 640 × 480 having 8 bits per pixel and 500 dpi stored either as a TIFF image or a BMP image is considered to be a good biometric image. Figure 2.2 shows some samples of raw fingerprint and iris images.

2.2.2.2 Feature Sets

Real World Applications of biometric systems require quick access to data and very small storage space, therefore, the feature sets are usually stored instead of raw images. Processed feature sets usually take up less space than raw images, thus

Fig. 2.2 A visual spectrum
fingerprint image, a thermal
fingerprint image and an IR
spectrum iris image
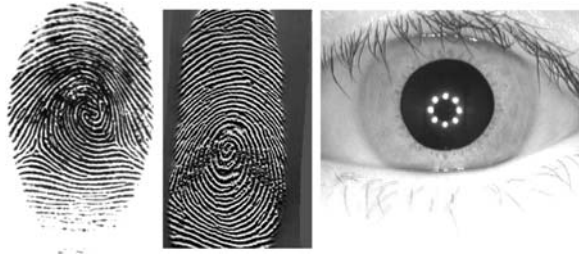
reducing the database size and also increasing the access speed. One of the lesser
known advantages of using feature sets is that it is not possible to recreate the actual
raw biometric data from the feature set therefore saving the feature set only provides
personal data protection.

It should be kept in mind that to maintain system openness, the feature sets should
be stored in one of the standard formats like the ones defined in ANSI/NBS – ICST
1-1986 for minutiae, ANSI/NIST – ITL 1a-1997 for Faisal Feature Set, ANSI/NIST
– ITL 1-2006 for Iris, etc. [4, 5]. Using these standards allows for an easy expansion
and upgrade of the system at later times.

2.3 Feature Extraction

As mentioned in Chapter 1, one of the first steps in feature extraction process is
pre-processing. Pre-processing can consist of multiple steps e.g.

Image Enhancement – To reduce the noise in the image and/or to enhance the
features to be extracted.

Image Formatting – Image is converted to a format that will allow for better
feature extraction performances; e.g. some minutiae extraction algorithms require
the image to be converted into a binary form.

Image Registration and Alignment – Images are aligned and centred so that the
extracted features from all images are within the same frame of reference thus mak-
ing the matching process much simpler.

Image Segmentation – Raw image is segmented into Region of Interest (ROI) and
background. All the processing is carried out on the ROI, it is therefore imperative
that the best possible segmentation is obtained.

Once a raw image is properly processed the feature extraction algorithm
can then be used to extract the relevant features. Feature Extraction Algorithms
can be classified into two groups: Global Feature Extractors and Local Feature
Extractors.

Global Feature Extractors aim to locate usable features from the raw data at the
overall image level. They process the image as a whole and try to extract the fea-
tures, e.g. Gabor Wavelets-based approach for iris and fingerprint recognition [6].

On the other hand, Local Feature Extractors focus on the chunks of image data. These algorithms work on small windows within the images and extract the relevant features, e.g. minutiae extraction from skeletonised and binarised fingerprint images.

A feature extractor algorithm selection is governed mainly by the type of application that the system is being designed for. Applications requiring more accuracy and security should have a robust and exhaustive feature extractor. However, for faster applications a simpler algorithm might be the best option. Ideally, the feature extractor should be very robust, accurate and fast but practically this is not possible. It is therefore almost always a compromise between accuracy and speed. It is advisable to evaluate multiple feature extraction algorithms to find the optimal algorithm for the desired application.

The feature extractor algorithm selection also depends upon the type of matcher being used in the system. The feature extractor should generate output in the format that the matcher is able to comprehend and process.

As mentioned before, to maintain openness of the system it is prudent to ensure that the output of the feature extractor should follow a standard format.

2.4 Matcher

A matcher algorithm takes the reference feature set and compares it with all the template feature sets in the database to provide a matching score for each pair. It then selects the best template–reference pair and outputs the details as its decision.

Different types of matchers are usually used depending upon the type and format of the feature set as well as the type of application at hand. Matchers are commonly categorised into two categories: Time Domain Matchers and Frequency Domain Matchers [1–3].

Time Domain Matchers work in the spatial domain and the feature sets for these types of matchers are generated directly from the raw images.

Frequency Domain Matchers operate in the frequency domain and the feature sets for these types of matchers are generated by first transforming the image into the frequency domain and then selecting the features, e.g. wavelets-based matchers, Fourier transform-based matchers, etc.

However, it is worth noting that correlation-based matchers are the most commonly used matcher algorithms. Similarly, distance-based matchers and supervised learning or pattern recognition-based matching is also widely used.

Pattern recognition-based matching finds the correct match by training on known correct and incorrect matching feature sets. In this type of matching the training process is usually computationally intensive, but if this process is efficiently done, matching can work very fast and provide highly accurate results.

The selection of the optimal matcher depends upon the application for which the system is being developed as well as the type of feature sets available for matching. In addition, information regarding the desired accuracy and the speed required also plays an important role when selecting a matcher algorithm.

2.5 System Testing

When a biometric identification system is developed it should be thoroughly tested before going live. Appropriate testing is critical to locate, identify and eradicate errors before the system is implemented in a real world scenario.

Testing is conducted on a test data set and for impartial testing of biometric recognition systems the industry and research community need to have access to large public data sets. Test data set is generated by collecting data with known values (called the ground truth) and the testing is carried out by matching a set of known reference feature sets with the test database and evaluating the result of the matcher against the ground truth [7]. Evaluation of testing is usually conducted in two overlapping categories: technology and operation scenarios. Technology evaluation is used to measure the performance of the recognition algorithm and is usually conducted using standard data sets. In general, the results from this type of evaluation are used to further improve the algorithm. On the other hand, the aim of the operation evaluation is to measure and assess the performance of the recognition system in a particular application; for example iris authentication at an airport. This type of evaluation usually takes weeks/months to account for environment changes including varying test samples.

The results are then analysed to locate the bugs and remove them. The performance of the system is also evaluated and presented in a standard format for the users.

2.6 Performance Evaluation

Determining the best biometric system for a specific operational environment and how to set up that system for optimal performance requires an understanding of the evaluation methodologies and statistics used in the biometrics community. The degree of similarity between two biometric images is usually measured by a similarity score. The similarity score is called genuine score if the similarity is measured between the feature sets of the same user. On the other hand, it is called imposter score if it is between feature sets of different users.

An end-user is often interested in determining the performance of the biometric system for *his specific application*. For example, he/she would like to know whether the system makes accurate identification. Although, there exists a few criteria, no metric is sufficiently adequate to give a reliable and convincing indication of the identification accuracy of a biometric system. However, one criterion generally accepted by biometric community uses either a *genuine individual* type of decision or an *impostor* type of decision, which can be represented by two statistical distributions called genuine distribution and impostor distribution, respectively. The performance of a biometric identification system can then be evaluated based on resulting Genuine Score and Imposter Score generated by the system [7]. The following are usually employed:

Matcher accuracy – Accuracy is measured on the test data set to discover how many of the feature sets are correctly matched by the system. If the genuine score is above an operating threshold of the system, the feature set is considered to be correctly matched. Matcher accuracy is usually displayed as a percentage of matches.

False accept rate (FAR) – If an imposter score is above the operating threshold it is called a false accept. FAR, therefore, means that the system accepted an imposter as a genuine user. FAR is one of the major performance matrixes that have to be closely evaluated. In fact, effort should be made to keep it as close to zero as possible.

False reject rate (FRR) – If a genuine score is below the threshold then it is called a false reject. Thus, false reject rate means that the system rejected a genuine user as an imposter. FRR should ideally be as close to zero as possible but in most access control applications it is not as critical as FAR. If a user is rejected as an imposter he/she can always try again but if an imposter is accepted as a genuine user the integrity of the complete system is compromised.

False alarm rate – A statistic used to measure biometric performance when operating in the watch-list (sometimes referred to as open-set identification) task. This is the percentage of times an alarm is incorrectly sounded on an individual who is not in the biometric system's database (the system alarms on John when John isn't in the database) or an alarm is sounded but the wrong person is identified (the system alarms on Peter when Peter is in the database, but the system thinks Peter is Daniel).

Equal error rate (ERR) – It is the point on the ROC curve where FAR and FRR are equal. For a high-performance system ERR should be as low as possible.

Most vendors provide the performance evaluation in terms of accuracy and ERR. Some other evaluation criteria are

Failure to capture rate (FCR) – FCR pertains to the amount of times a sensor is unable to capture an image when a biometric trait is presented to it. The FCR increases with wear and tear to the sensor module. If the FCR increases above a certain threshold it is advisable to replace the sensor module.

Failure to enrol (FTE) – FTE indicates the number of users that were not enrolled in the system. FTE is usually related to the quality of the biometric image. In most cases, a system is trained to reject poor quality images. This helps in improving the accuracy of the system and reducing the FAR and FRR. Every time an image is rejected the FTE is increased. A trade-off between quality and FTE is required if the system is to be accepted by the users.

2.7 Conclusion

To develop a strong biometric system it is imperative to select a very stable data acquisition system and a very secure, fast and robust database. Feature Extractor and Matcher selection will directly impact the user acceptance of the system and the selection is based on the type of application.

References

1. Arun A. Ross, Patrick Flynn and Anil K. Jain, "Handbook of Biometrics" ISBN: 978-0-387-71040-2.
2. A. K. Jain, A. Ross and S. Prabhakar, "An introduction to biometric recognition," IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no. 1, January 2004.
3. J. G. Daugman, "Biometric Decision Landscape", Technical Report No. TR482. University of Cambridge Computer Laboratory, 1999.
4. Data Format for Information Interchange – Fingerprint Identification, ANSI/NBS – ICST 1-1986.
5. Data Format for Information Interchange – Data Format for the Interchange of Fingerprint, Facial & SMT Information, ANSI/NIST – ITL 1a-1997.
6. H. Meng and C. Xu, "Iris Recognition Algorithm Based on Gabor Wavelet Transform," IEEE International Conference on Mechatronics and Automation, 2006.
7. J. Wayman, A. Jain, D. Maltoni and D. Maio, "Biometric Systems Technology, Design and Performance Evaluation," ISBN: 1852335963.