12

Human Recognition using Face, Fingerprint and Voice

We describe in this chapter a new approach for human recognition using as information the face, fingerprint, and voice of a person. We have described in the previous chapters the use of intelligent techniques for achieving face recognition, fingerprint recognition, and voice identification. Now in this chapter we are considering the integration of these three biometric measures to improve the accuracy of human recognition. The new approach will integrate the information from three main modules, one for each of the three biometric measures. The new approach consists in a modular architecture that contains three basic modules: face, fingerprint, and voice. The final decision is based on the results of the three modules and uses fuzzy logic to take into account the uncertainty of the outputs of the modules.

12.1 Introduction

The term "biometrics" is derived from the Greek words bio (life) and metric (to measure). We typically choose to interpret biometrics as methods for determining unique (or relatively unique, if such an expression is allowed) features of a person's body to distinguish them from the rest of humanity. Note that there is also a branch of statistics with the same name. This branch deals with all types of data pertaining to variability in human form, some of which is of value to the "biometrics" of the computer industry.

The concept of identification based on aspects of the human body is certainly not a new one. As early as the 14th century, the Chinese were reportedly using fingerprints as a form of signature. During the late 1890s, a method of bodily measurement, called Bertillonage, (after its founder Alphonse Bertillon), was used by police authorities throughout the world. This system quickly faded when a case of two indistinguishable people with almost identical names was discovered. From this point on, fingerprinting

Patricia Melin and Oscar Castillo: *Hybrid Intelligent Systems for Pattern Recognition Using Soft Computing*, StudFuzz **172**, 241–256 (2005)
www.springerlink.com

(developed by Richard Edward Henry of Scotland Yard) became essentially the only identification tool for police.

Today, a variety of methods and techniques are available to determine unique identity, the most common being fingerprint, voice, face, and iris recognition. Of these, fingerprint and iris offer a very high level of certainty as to a person's identity, while the others are less exact. A large number of other techniques are currently being examined for suitability as identity determinants. These include (but are not limited to) retina, gait (walking style), typing style, body odour, signature, hand geometry, and DNA. Some wildly esoteric methods are also under development, such as ear structure, thermal imaging of the face and other parts of the body, subcutaneous vein patterns, blood chemistry, anti-body signatures, and heart rhythm, to name a few.

12.1.1 Major Biometric Methods

The four primary methods of biometric authentication in widespread use today are face, voice, fingerprint, and iris recognition. All of these are mentioned in this chapter, some more abundantly than others. Generally, face and voice are considered to be a lower level of security than fingerprint and iris, but on the other hand, they have a lower cost of entry. We describe briefly in this section some of these biometric methods.

Face Recognition

Facial recognition has advanced considerably in the last 10 to 15 years. Early systems, based entirely on simple geometry of key facial reference points, have given way to more advanced mathematically-based analyses such as Local Feature Analysis and Eigenface evaluation. These have been extended though the addition of "learning" systems, particularly neural networks.

The task of identifying people by face recognition can be easily divided into two sub-tasks. Firstly, there is the task of recognizing a face in a scene and extracting it from the surrounding "noise. Secondly, any computer system must then be able to extract unique features from that image and compare them with images already stored, hopefully correctly identifying the subject at hand. Based on the source of the facial image, there may also be a need to perform "liveness" tests such as determining that the head is moving against the background and that a degree of three-dimensionality can be observed. Other systems are intended to work from previously recorded images and don't test for these factors.

Face recognition systems are particularly susceptible to changes in lighting systems. For example, strong illumination from the side will present a vastly different image to a camera than neutral, evenly-positioned fluorescent lighting. Beyond this, however, these systems are relatively immune to

changes such as weight gain, spectacles, beards and moustaches, and so on. Most manufacturers of face recognition systems claim false accept and false reject rates of 1% or better.

In our hybrid intelligent approach, a typical face-recognition system would rely on a simple Web camera to acquire an image prior to authentication processing, with the user gently moving his or her head while the software captures one or more images. PC resource requirements are no more than for a typical desktop PC. In an environment where moderate security is sufficient and there is no desire to purchase additional hardware (assuming the Web cameras are available), this is a very useful authentication technique.

Voice Recognition

Software systems are rapidly becoming adept at recognizing and converting free-flowing speech to its written form. The underlying difficulty in doing this is to flatten out any differences between speakers and understand everyone universally. Alternatively, when the goal is to specifically identify one person in a large group by their voice alone, these very same differences need to be identified and enhanced.

As a means of authentication, voice recognition usually takes the form of speaking a previously-enrolled phrase into a computer microphone and allowing the computer to analyze and compare the two sound samples. Methods of performing this analysis vary widely between developers. None is willing to offer more than cursory descriptions of their algorithms–principally because, apart from LAN authentication, the largest market for speaker authentication is in verification of persons over the telephone.

A number of issues contrive to weaken the performance of voice recognition as a form of biometric verification of identity. For instance, no two microphones perform identically, so the system must be flexible enough to cope with voiceprints of varying quality from a wide range of microphone performance. Also, the same person will not speak the code-phrase the same from one day to the next, or even from one minute to the next. Most researchers claim false accept and false reject rates of around 1–2%, although our research work suggests this may be excessively confident. It is safe to assume that every desktop PC in an organization is either already fitted with a microphone or can easily have one added at minimal cost. This is the main attraction of voice-based identification schemes–no other equipment or significant resources are required.

Fingerprint Recognition

The process of authenticating people based on their fingerprints can be divided into three distinct tasks. First, you must collect an image of a fingerprint; second, you must determine the key elements of the fingerprint for confirmation

of identity; and third, the set of identified features must be compared with a previously-enrolled set for authentication. The system should never expect to see a complete 1:1 match between these two sets of data. In general, you could expect to couple any collection device with any algorithm, although in practice most vendors offer proprietary, linked solutions.

A number of fingerprint image collection techniques have been developed. The earliest method developed was optical: using a camera-like device to collect a high-resolution image of a fingerprint. Later developments turned to silicon-based sensors to collect an impression by a number of methods, including surface capacitance, thermal imaging, pseudo-optical on silicon, and electronic field imaging.

Unlike with the previous two methods (voice and face), it is unlikely that a typical desktop PC will be equipped with fingerprint-capture hardware. Most fingerprint reader units connect to either the parallel port (and piggy-back on the PS/2 for power) or make a USB connection. In addition to standalone units, there are fingerprint readers mounded in keyboards and in combination with smart-card readers. Recently, several laptop manufacturers have begun providing a capacitance-based fingerprint scanner either mounted next to the keyboard or as a PCMCIA-attachable unit.

The process of confirming identity via fingerprint is quick and simple for the user. Upon activation of the authentication request, most systems will energize the reader and place a viewing window on the screen. The user places a finger on the reader and can observe the quality of the captured image on the viewing window. Most systems will automatically proceed to the analysis phase upon capture of a good-quality fingerprint, but some require the user to press the <Enter> key.

The most processor-intensive portion of the recognition sequence is analyzing the scanned image to determine the location of ridges and subsequently the identification of points of termination and joining. For this phase, most manufacturers suggest a PC with a Pentium processor of at least 120 MIPS. Recent developments in smart card technology have allowed sufficient processing power to perform the actual template match on the card. This means that an enrolled template can be stored on the "hidden" side of a crypto-card and need never be released outside the card–an important step in promoting the privacy of biometric templates. However, a PC is still required to perform the image analysis.

As discussed, a variety of fingerprint detection and analysis methods exist, each with their own strengths and weaknesses. Consequently, researchers vary widely on their claimed (and achieved) false accept and false reject rates. The poorest systems offer a false accept rate of around 1:1,000, while the best are approaching 1:1,000,000. False reject rates for the same vendors are around 1:100 to 1:1000.

It is generally accepted that fingerprint recognition systems offer a moderately-priced solution with very good abilities to accurately confirm user

identity. For this reason, this is the most widely-used biometric method in office environments.

Iris Recognition

Iris recognition is based entirely on a concept originated by Drs. Leonard Flom and Aran Safir, and a software process developed by Dr. John Daugman, all of Cambridge University, England. US Patent 5,291,560 issued in the name of Daugman has been assigned to Iridian Corp., one of the world's principal companies of iris-based systems. Extensive research has determined that the human iris is essentially unchanged in structure and appearance from the eighth month of gestation until a few minutes after death. Although a neonatal eye can be darkly coloured until a few months after birth, this darkness is not an influence in the infrared wavelengths normally used to collect an iris image.

Following identification and extraction of the iris from a captured image of the eye, a pseudo-polar coordinate system is established over the iris image. This allows a large number of two-dimensional modulation waveforms to be extracted that are invariant of changes due to iris widening (due to light levels) or to external factors such as spectacles or contact lenses. Iris recognition systems require special iris cameras with very high resolution and infrared illumination abilities. Typically these are attached to PCs via USB connectors. The user is required to look squarely into the camera while the eye is illuminated with a focussed infrared source. Elapsed time from capture to confirmation of identity takes less than a second. Performance requirements are no greater than those already available on the typical desktop PC.

In the history of iris recognition, there has *never* been a false acceptance. In fact, the equal error rate is 1:1,200,000, with a typical false accept rate of 1:100,000,000 and false reject rate of 1:200,000. Note that these are theoretical values based on strong analysis of limited data (only 5 to 10 million iris scans have ever been performed); they also do not take into account the perceived level of difficulty in using the system. Overall, iris scanning is the system to use if you are concerned about strongly authenticating users. The devices are considerably more expensive than fingerprint readers, but the gain in authentication confidence more than offsets the increased cost.

12.2 Biometric Systems

We describe in this section some basic concepts about biometric systems, as well as methods to evaluate their performance. We also describe the basic components of a biometric system and comparison of the different biometric measures.

Table 12.1. General comparison of biometric measures

Biometric Type	Accuracy	Ease of Use	User Acceptance
Fingerprint	High	Medium	Low
Hand Geometry	Medium	High	Medium
Voice	Medium	High	High
Retina	High	Low	Low
Iris	Medium	Medium	Medium
Signature	Medium	Medium	High
Face	Low	High	High

12.2.1 Types of Biometrics

Several different biometric modalities have emerged in recent years. The Table 12.1 lists the more common biometric sources of identity information and key characteristics of some current systems; classified in broad terms:

It is important to note that some techniques, such as retinal scanning or finger print recognition, may offer high accuracy but may not be appropriate for some applications. This is due to the high level of cooperation required by the user or the social or psychological factors that may prove unacceptable to potential users.

Both voice and face recognition are considered to be easy to use and normally acceptable by potential users. However, their accuracy is currently less than some other biometric technologies, especially in unconstrained environments such as where the background sound and illumination is variable. More information on the characteristics of specific biometric modalities can be found in (Jennings, 1992).

12.2.2 Biometric System Components and Processes

There are two distinct phases of operation for biometric systems: enrolment and verification/identification. In the first phase identity information from users is added to the system. In the second phase live biometric information from users is compared with stored records. Typical biometric identification and recognition system may have the following components:

(a) Capture: A sub-system for capturing samples of the biometric(s) to be used. This could be voice recordings or still facial images. Specific features will be extracted from the biometric samples to form templates for future comparisons. In the enrolment phase a number of such samples may be captured. A truly representative identity model may then be obtained from the features thus obtained. This enrolment process should ideally be simple and rapid, yet result in, good quality, representative templates. If the templates are of poor quality, this will affect the subsequent performance of the system. An elaborate and exhaustive enrolment process may be unacceptable.

(b) Storage: The templates thus obtained will have to be stored for future comparison. This may be done at the biometric capture device or remotely in a central server accessible via a network. Another alternative is to store the template in a portable token such as a smart card. Each one of these options has its advantages and disadvantages (Ashbourn, 1999). In addition to template storage there is often a need for a secure audit trail for all the transactions of the system.

(c) Comparison: If the biometric system is used in a verification setting, then the claimed user identity will have to be compared against the claimed reference template. The captured live biometric from the user will be compared with the claimed identity which may be provided by entering a pin, or presenting a card storing identity information. In a identification/recognition setting the live biometric will have to be compared with all the templates stored to see if there is a close match. In some systems it may be possible to automatically update the reference template after each valid match. This will make it possible for the system to adapt to gradual minor changes in user characteristics (e.g. due to aging).

(d) Interconnections: There is the need for interconnections between the capture device and the verification and storage components of the system. Often there are existing access control and information systems into which the biometric system may have to be integrated. There is a need for generic networking and programming interfaces to allow easy interconnections for biometric systems. Security and efficiency will be key considerations.

12.2.3 Biometrics System Considerations

The following are some of the key issues that need to be considered in designing and applying biometric systems.

Robustness

It is important to consider how robust the system is to fraud and impersonation. Such fraud can occur at the enrolment stage as well as at the verification stage. Using more than one biometric modality can help combat fraud and increase robustness. Also the system should be robust to small variations to the users' biometrics over time. For this, an adaptive system that gradually modifies the stored templates may be used.

Acceptability: The technology must be easy to use during both the enrolment and comparison phases. It must also be socially acceptable. The users would not accept a system that may threaten their privacy and confidentiality or that might appear to treat them as potential suspects and criminals. This accounts for the lower acceptability of fingerprint systems than voice or face recognition systems. A multimodal system is more capable to adapting to user's requirements and capabilities.

Legal issues may also have to be considered in relation to biometric systems (Woodward, 1997). There may be concerns over potential intrusions into private lives by using biometric systems. The European Union's comprehensive privacy legislation, the Directive on Data Protection, became effective on October 25, 1998. While it does not specifically mention biometrics, biometric identifiers are likely to fall within its legislative scope. The European Parliament has recently raised this issue in relation to European Community research efforts. Also, there is a growing lobby to limit and regulate the use of biometrics and surveillance technologies. Legal issues must be considered for any potential application and appropriate measures must be taken. A clear public stance on the issue of privacy in relation to biometric technologies is required to ensure broad public acceptance.

Speed and Storage Requirements

The time required to enroll, verify or identify a person is of critical importance to the acceptance and applicability of the system. Ideally, the acceptable verification time should be of the order of one second or faster. The storage requirement for the templates is also an important issue, especially if the templates are to be stored in magnetic stripe or smart cards.

Integration

The hardware platform on which the system is to be implemented is a key concern. The software, hardware and networking requirements should ideally be compatible with existing systems, allowing the biometric system to be integrated to the existing infrastructure. The system cost should be reasonable and the maintenance costs should be understood.

12.2.4 Performance Assessment

An important issue for the adoption of biometric technologies is to establish the performance of individual biometric modalities and overall systems in a credible and objective way.

For verification applications, a number of objective performance measures have been used to characterize the performance of biometric systems. In these applications a number of "clients" are enrolled onto the system. An "impostor" is defined as someone who is claiming to be someone else. The impostor may be someone who is not enrolled at all or someone who tries to claim the identity of someone else either intentionally or otherwise. When being verified the clients should be recognized as themselves and impostors should be rejected.

False Acceptance Rate (FAR) is defined as the ratio of impostors that were falsely accepted over the total number of impostors tested described as a percentage. This indicates the likelihood that an impostor may be falsely accepted and must be minimized in high-security applications.

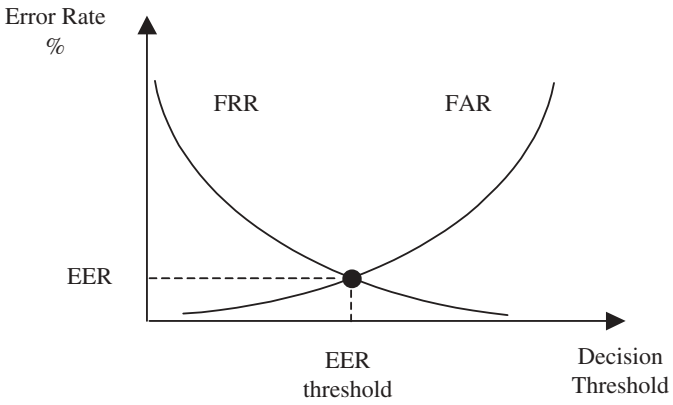

Fig. 12.1. FRR and FAR curves

False Reject Rate (FRR) is defined as the ratio of clients that are falsely rejected to the total number of clients tested described as a percentage. This indicates the probability that a valid user may be rejected by the system. Ideally this should also be minimized especially when the user community may be put-off from using the system if they are wrongly denied access.

The biometric verification process involves computing a distance between the stored template and the live sample. The decision to accept or reject is based on a pre-defined threshold. If the distance is less than this threshold then we can accept the sample. It is therefore clear that the performance of the system critically depends on the choice of this threshold and there is a trade-off between FRR and FAR. Vendors usually provide a means for controlling the threshold for their system in order to control the trade-off between FRR and FAR. The Equal Error Rate (EER) is the threshold level for which the FAR and the FRR are equal. Figure 12.1 shows a general example of the FRR and FAR curves. The EER is often quoted as a single figure to characterize the overall performance of biometric systems. Another important performance parameter is the verification time defined as the average time taken for the verification process. This may include the time taken to present the live sample.

The EU funded BIOTEST project is one initiative to provide objective performance characterization of biometric products. A National Biometric Test Centre has been established in the US and similar efforts are underway in other countries. A number of databases have been developed for the evaluation of biometric systems (Chibelushi, 1996). For the testing of joint audio-visual systems a number of research databases have been gathered in recent years (Messer et al., 1999). Developing new assessment strategies that allow meaningful comparisons between systems and solutions is an essential activity. This involves creating databases and putting together test procedures and systems for the online assessment of biometric technologies.

12.3 Architecture for Human Recognition

Our proposed approach for human recognition consists in integrating the information of the three main biometric parts of the person: the voice, the face, and the fingerprint. Basically, we have an independent system for recognizing a person from each of its biometric information (voice, face, and fingerprint), and at the end we have an integration unit to make a final decision based on the results from each of the modules. In Fig. 12.2 we show the general architecture of our approach in which it is clearly seen that we have one module for voice, one module for face recognition, and one module for fingerprint recognition. At the top, we have the decision unit integrating the results from the three modules.

As we have described in the previous three chapters the recognition systems for the face, fingerprint and voice of a human person, now we only need to concentrate in describing how to integrate the results from these three modules. The decision unit at the top of the hierarchy in Fig. 12.2 is the one that will integrate the results from the three modules. This decision unit uses fuzzy logic to take into account the uncertainty involved in the decision process.

We use, in the decision unit, a set of fuzzy rules to make final decision of human recognition. The fuzzy system has three input linguistic variables, which are face, fingerprint, and voice. Since each of these variables will have the result of the corresponding module with certain level of uncertainty, the fuzzy rules will take into account the values of the variables to give the final

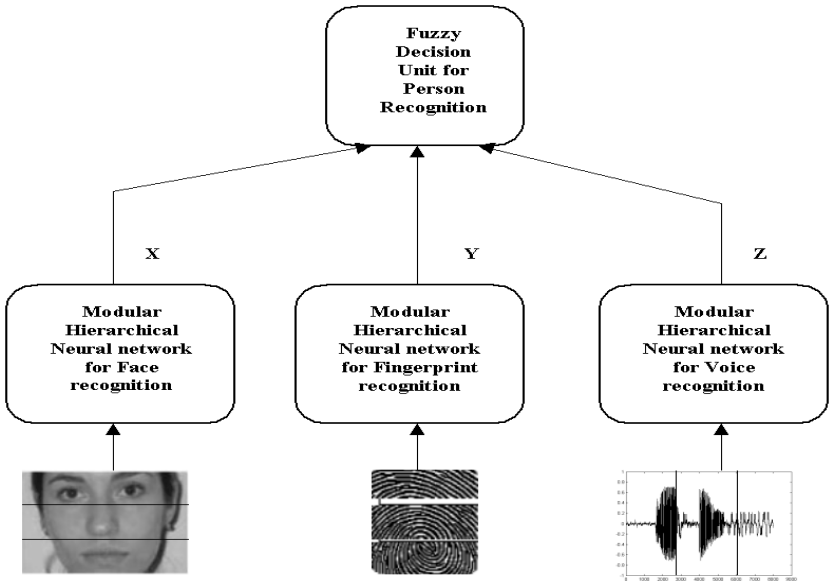

Fig. 12.2. Architecture of the proposed approach

output, which will be a specific identification of the person. We will describe the fuzzy system for decision in the following section.

12.4 Fuzzy System for the Decision on Human Recognition

We describe in this section a fuzzy system to integrate the outputs of three modules of the human recognition system. The linguistic variables of the fuzzy system are: Face, Fingerprint, and Voice. We will assume that X, Y, and Z, are three possible identifications of persons in the database. We need only to consider three possible values because we have three modules and in the worst case we will have three different results. Of course, the easiest case is when we have positive agreement in the identification of all the variables, which is the case illustrated in rule 1. The other easy case is when we have negative agreement in the three variables, which is the case illustrated in rule 9. For other cases, when we have two values in agreement and the third one is different, the output will be the majority value. When the three values are different, then the output will depend on the highest membership. Also, we can take into account that the fingerprint recognition is more reliable (according to Table 12.1), and then the voice recognition. Of course, face recognition is the least reliable of three methods used. The fuzzy rules are given as follows:

Rule 1: IF Face = X AND Fingerprint = X AND Voice = X
THEN Person = X
Rule 2: IF Face = X AND Fingerprint = X AND Voice = Y
THEN Person = X
Rule 3: IF Face = X AND Fingerprint = Y AND Voice = X
THEN Person = X
Rule 4: IF Face = Y AND Fingerprint = X AND Voice = X
THEN Person = X
Rule 5: IF Face = X AND Fingerprint = Y AND Voice = Y
THEN Person = Y
Rule 6: IF Face = Y AND Fingerprint = Y AND Voice = X
THEN Person = Y
Rule 7: IF Face = Y AND Fingerprint = X AND Voice = Y
THEN Person = Y
Rule 8: IF Face = X AND Fingerprint = Y AND Voice = Z
THEN Person = Y
Rule 9: IF Face = Z AND Fingerprint = Z AND Voice = Z
THEN Person = Z

We now describe an implementation of this fuzzy system for verifying and validating the results of the approach for integrating the decisions of the three modules. First, we show in Fig. 12.3 the architecture of the fuzzy system,

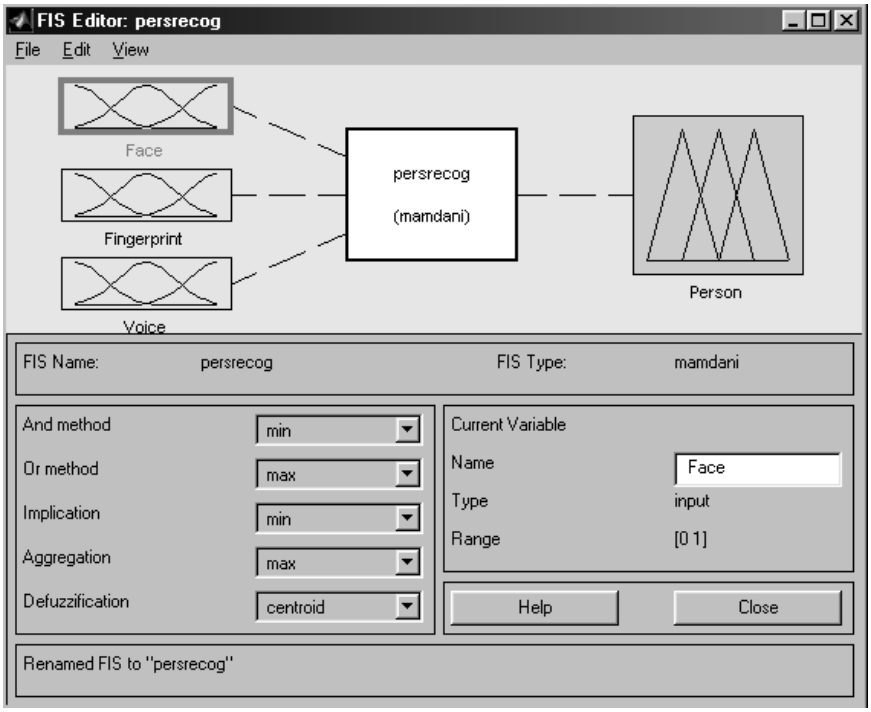

Fig. 12.3. Architecture of the fuzzy system for person recognition

which has three input linguistic variables and one output linguistic variable. We are using a Mamdani type fuzzy model with the max-min inference method and centroid deffuzzification. We also show in Figs. 12.4, 12.5, 12.6, and 12.7 the membership functions of all the linguistic variables involved in the fuzzy system. We have to note that all the membership functions are Gaussian. Regarding the membership values used in the inference, these values come from the degree of certainty of the decisions of face, fingerprint, and voice. In each case, when a decision is reached in the respective module, the final result has a degree of certainty between 0 and 1. This value is used as a membership degree in the respective linguistic value of the variable. We show in Figs. 12.8 and 12.9 two cases of identification with specific input values, which are representative of the fuzzy system use. Finally, we show in Fig. 12.10 the non-linear surface representing the fuzzy model of person recognition.

We performed extensive tests on this fuzzy system for person recognition with a database of 100 individuals from our institution (with different levels of noise, up to 100%) and the recognition rate was of about 99%, which is acceptable. Still, we can fine-tune this fuzzy system with more data, or we can improve uncertainty management by using type-2 fuzzy logic or intuitionistic fuzzy logic. We will consider these two options as future research work.

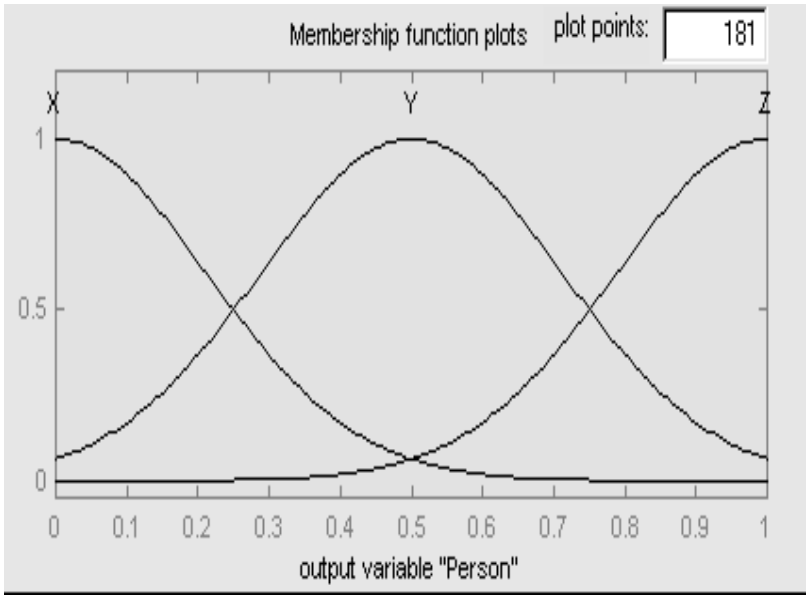

Fig. 12.4. Membership functions for the "person" linguistic output variable

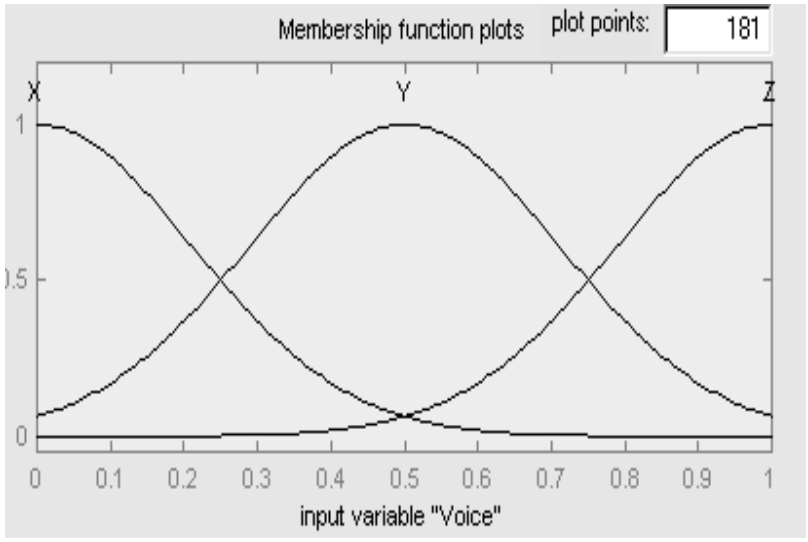

Fig. 12.5. Membership functions for the "voice" input linguistic variable

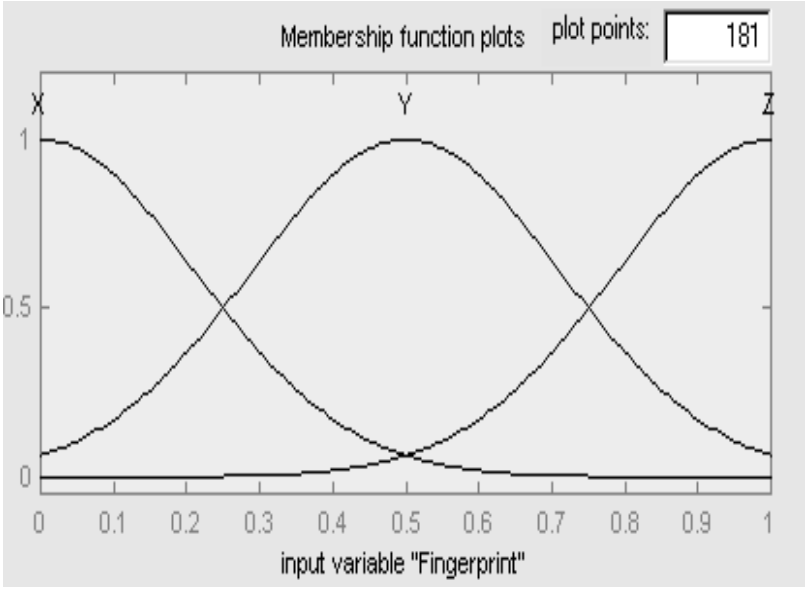

Fig. 12.6. Membership functions for the "fingerprint" input linguistic variable

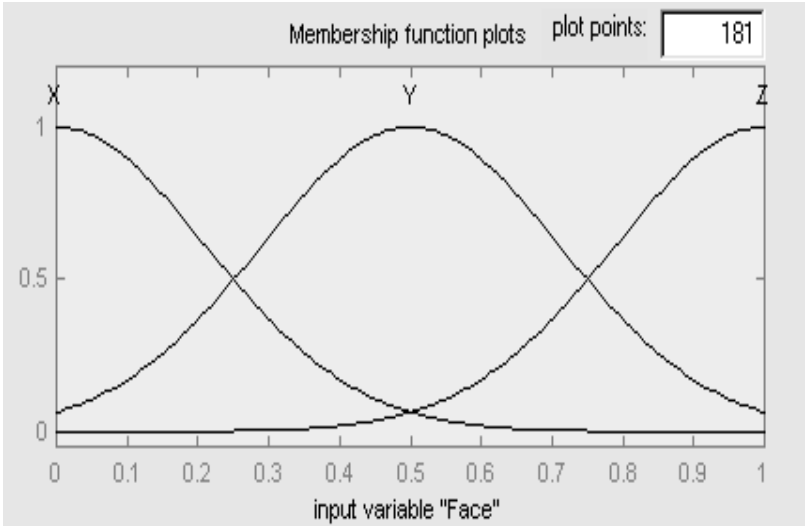

Fig. 12.7. Membership functions for the "face" linguistic variable

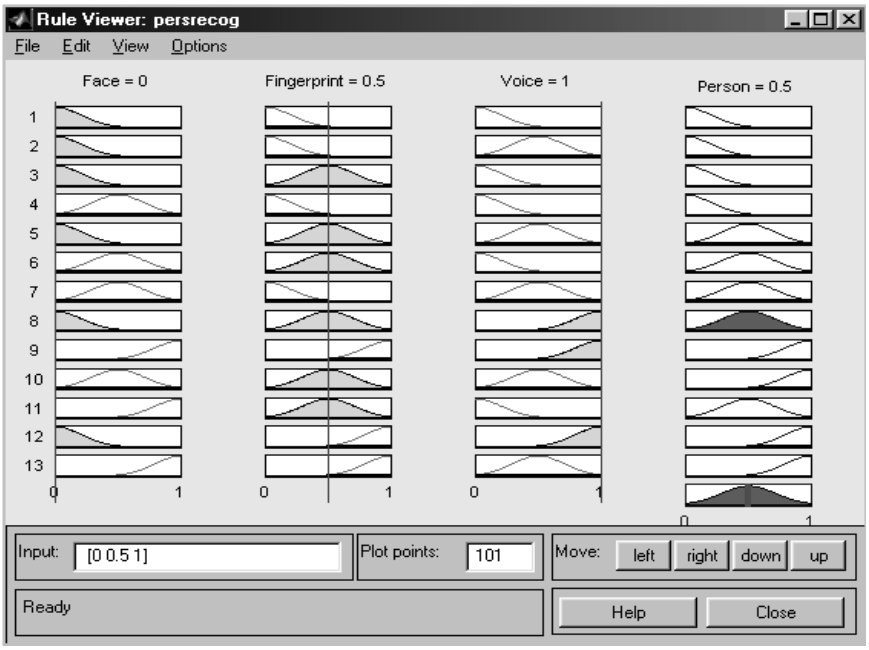

Fig. 12.8. Identification of person Y with specific input values

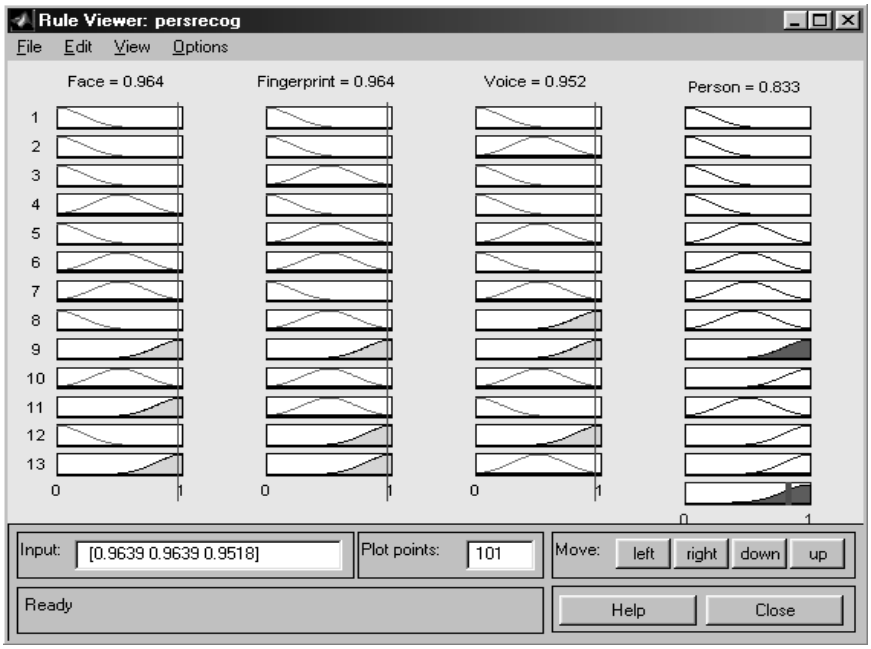

Fig. 12.9. Identification of person Z with specific input values

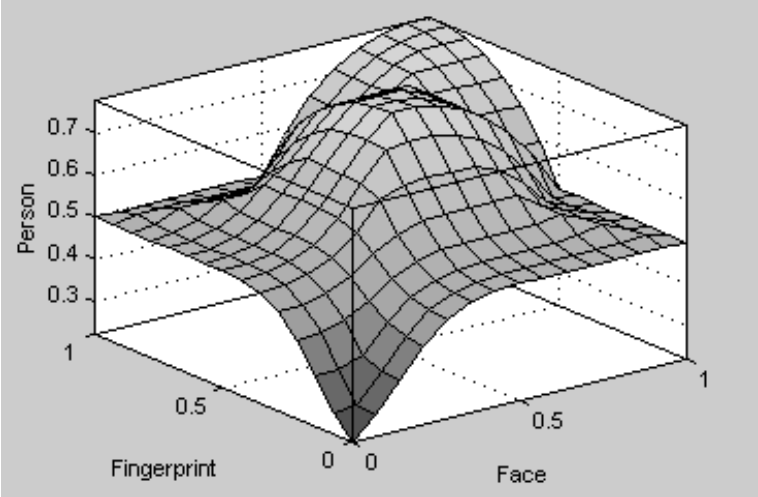

Fig. 12.10. Non-linear surface representing the fuzzy model of person identification

12.5 Summary

We described in this chapter our intelligent approach for integrating the results of face, fingerprint and voice recognition. The proposed approach consists in the use of a fuzzy system to implement the decision unit of the hierarchical architecture of human recognition. The fuzzy system consists of a set of fuzzy rules, which take into account the decisions of the individual modules of face, fingerprint and voice. The output of the fuzzy rules is the final identification of the person based on the input values of the three modules. We have achieved excellent results with this fuzzy logic approach for integrating the decisions of face, fingerprint and voice.