Chapter 13

Developing a Multimodal Biometric Authentication System Using Soft Computing Methods

Mario Malcangi

Abstract

Robust personal authentication is becoming ever more important in computer-based applications. Among a variety of methods, biometric offers several advantages, mainly in embedded system applications. Hard and soft multi-biometric, combined with hard and soft computing methods, can be applied to improve the personal authentication process and to generalize the applicability.

This chapter describes the embedded implementation of a multi-biometric (voiceprint and fingerprint) multimodal identification system based on hard computing methods (DSP) for feature extraction and matching, an artificial neural network (ANN) for soft feature pattern matching, and a fuzzy logic engine (FLE) for data fusion and decision.

Key words Artificial neural network, Fuzzy decision logic, Soft-biometric data, Multi-biometric

1 Introduction

More and more of the devices we deal with every day depend on embedded systems. This trend can reasonably be expected to accelerate going forward. Because of this, as well as due to other factors, the security of systems and of data is likely to continue to present challenges that developers will attempt to meet by devising new tools. There can be little doubt that many such methods will rely on the biometric approach [1–4].

Biometrics is uniquely suited to authentication tasks in embedded systems because it offers simultaneous solutions to the twin problems of the meager demand for resources that we wish to place on such systems and of the high security requirements that their potential applications can be expected to mandate [5, 6]. It is not hard to imagine wanting to limit demand on resources if we start from the very basic concept of a system that has no keyboard, is designed to have the smallest possible footprint, and can be considered subsidiary to a larger system, machine, or process. If we consider the embedded system's role as gatekeeper to a much

Hugh Cartwright (ed.), *Artificial Neural Networks*, Methods in Molecular Biology, vol. 1260,
DOI 10.1007/978-1-4939-2239-0_13,

larger and more complex reality, say the electronic ignition key to a sophisticated, automated installation, it is not hard, either, to envision the degree of security requirements for which a minuscule device is ultimately responsible. As such devices proliferate, the demand for lightweight yet robust authentication methods can only grow.

Because biometrics depends on physiological traits or on distinctive behavior [7, 8] unique to the individual whose identity is to be authenticated, or—indeed—a combination of the two, it can rightly stake a claim to being superior, at least potentially, to current and established authentication methods like personal identification numbers (PINs), passwords, or smart cards. The individual's biometric data is always available, is nonrandomly unique, cannot be transferred to another party, cannot be forgotten, is not subject to theft, and cannot be guessed. These advantages mean that biometrics provides very high security compared to traditional identification methods that rely on the possession of a token, such as a card, key, or chip, or of personal knowledge, as in the case of a password [9].

Despite the superiority of biometric authentication due to such advantages and despite its rapid spread in microelectronics, mass-market adoption has lagged. The primary reason widespread application of biometrics has not taken off is that it does not offer surefire authentication the way a password does. A password can always perform. On the other hand, biometric features, in their original form, are analog information. As a result, they are subject to variation when captured by a biometric scanning device. This applies to fingerprint readers, microphones for voice-pattern recognition, cameras for facial feature matching, and any other digital system that has to match measured, analog, human traits. Of course such features can be digitized, but the data will nonetheless have been processed through fuzzy logic. Fuzziness can only be minimized so much, given that the original, analog features are, themselves, inherently fuzzy.

False acceptance rate (FAR) and false rejection rate (FRR) [10] can, of course, be minimized with classical pattern-matching techniques [11–13] but a 100 % correct acceptance rate cannot be achieved this way. However, the intrinsic fuzziness of biometric features makes it logical to look to soft-computing approaches [14–16] in the design of processes that match biometric feature patterns in the hope that nearly 100 % correct authentication might be attained that way. For example, even in prohibitive conditions, human beings always manage to recognize a familiar face. The human mind processes biometric features fuzzily and neurally.

Soft-computing data are variations and uncertainties. Biometric features vary constantly, present challenges to analytical description, and inherently fall short of belonging to their owner 100 % of the time.

It follows that soft-computing methods ought to work well for measuring and matching biometrics. Moreover, what humans do to reach nearly 100 % authentication rates amounts to acquiring data from multiple sources. They combine speech traits and facial features when matching a biometric identity to an individual [17]. In other words, they carry out neural and fuzzy processing on aggregated biometric feature sets.

While biometric authentication has seen surprisingly slow adoption generally, this has proved even more the case with the application of traditional biometric solutions to embedded systems [18]. Because embedded systems have limited resources—smaller memory and slower processors—they are ill suited to hungry authentication applications. Consequently, current installations of embedded systems typically rely on PIN codes and the like.

However, an authentication method based on soft biometric [19] data has the potential to be more miserly with computing resources than hard biometrics, using less data to map the identity of the person to be authenticated. Furthermore, multiple-criteria biometric authentication processes [8, 20, 21] can be devised so as to match the input capacities of embedded systems. By combining soft-computing methods and multi-biometrics, we can optimize such authentication for implementation on embedded systems.

One approach that holds promise in this regard involves setting up authentication based on a combination of voiceprint and fingerprint that uses hard-computing digital signal processing to extract features and match them but then turns to an artificial neural network (ANN) [22] for soft feature pattern matching and to a fuzzy logic inferential engine (FLE) for data fusion and decision making. Such a setup can be designed to provide highly robust, personal, biometric authentication. Experiments have been carried out where this can be accomplished with a single-chip, floating-point, digital signal processor.

2 Materials

To design a multimodal soft computing-based hard/soft biometric embedded system several methodologies and technologies occur:

- Biometric sensors.
- Sensor data acquisition.
- Feature extraction algorithms.
- Hard computing pattern matching.
- Soft computing pattern matching (artificial neural network).
- Soft computing fusion and decision (fuzzy logic).
- System modeling and simulation environments.

2.1 Biometric Sensors

The biometric sensors are specifically designed to capture physiologic or behavioral signals generated by human beings. The sensor is the first device of the signal chain. It captures and converts a physical property into analog signals (commonly electrical) or digital signals (when it embeds the mixed-signal electronics).

2.1.1 Voiceprint Sensors

The voiceprint sensor is a voice-grade microphone system able to capture the utterance, preserving the intelligibility of the speech.

- Single microphone: captures the acoustic wave of the utterance as a direct sound source (position sensitive; sensitive to surrounding noise).
- Dual microphone: captures the acoustic sound wave of the utterance and the surrounding audio noise (partially position sensitive; low sensitivity to surrounding noise).
- Microphone array: captures the acoustic sound wave of the utterance by a set of microphones spatially distributed (position insensitive; highly insensitive to surrounding noise).

A single microphone has been used in this design (*see* **Note 1**).

2.1.2 Fingerprint Sensors

A fingerprint sensor is an electronic device capable to capture the image of the fingerprint pattern and make it available as a bit map. Several technologies have been applied to develop fingerprint sensors:

- Optical (special digital camera): captures visible light emitted from a phosphor layer which illuminates the surface of the finger—advantages: noncontact, not electrostatic discharge sensitive—disadvantages: capabilities affected by the quality of skin, easily fooled.
- Ultrasonic (based on medical ultrasonic imaging principles): captures the images of a fingerprint by penetrating the epidermal layer of skin with high-frequency sound—advantages: noncontact, not electrostatic discharge sensitive, capabilities not affected by the quality of skin, very difficult to fool—disadvantages: the price is significantly higher than for a capacitive fingerprint scanner.
- Capacitance (based on electrical capacitance): consists of an array of capacitors in which one of the two plates is the dermal layer and the dielectric is the epidermal
  - Passive: Capacitance is measured across each sensor capacitor to form a pixel of the fingerprint image.
  - Active: A voltage is applied to the skin first, and then the charge of each capacitor of the array is measured to form a pixel of the fingerprint image.

Advantages: compatible with microelectronic integration—disadvantages: electrostatic discharge sensitive.

A passive capacitive sensor has been used in this design (*see* **Note 2**).

2.2 Sensor Data Acquisition Chain

Sensor data acquisition concerns the conditioning and digital representation of the voiceprint and fingerprint information captured by the respective biometric sensors. Conditioning is necessary if the sensor is affected by nonlinearities and if it is not dynamically compatible with analog-to-digital conversion (ADC) subsystem.

2.2.1 Microphone Sensor Data Acquisition

The microphone sensor data acquisition process is sensitive to several nonlinearities and mixed signal constraints. The microphone chain path needs conditioning and digital-to-analog adaptation, so optimal data can be available at the processing stage:

- Transducer linearization.
- Automatic gain control.
- Filtering.
- Sampling.
- Quantization.

The microphone nonlinearities need to be compensated to avoid distorted measurements at the level of feature extraction. This can be done in the analog domain by electronic circuitry tuned on the specific microphone, or in the digital domain at the preprocessing stage of the signal acquisition. In this design the second option has been applied because it is more flexible and adaptive. The microphone transfer function is estimated at calibration time, and then the nonlinearity compensated for by applying at run time the inverse function to the captured signal.

The amplification is required because the small signals from the microphone need to be adapted to the analog-to-digital converter (ADC) input dynamic range to have higher signal-to-quantization-noise ratio (SQNR).

Low-pass filtering (antialiasing) and high-pass filtering (offset removal) are applied prior to sampling the microphone signal to ensure lowest frequency distortion of the pulse code modulated (PCM) stream to be quantized and optimal SQNR.

2.2.2 Fingerprint Sensor Data Acquisition

The solid-state capacitive fingerprint sensor integrates all the required conditioning and digitalization resources so that optimal digital fingerprint image data is available at its output. Fingerprint image is captured at 513 dpi (dot per inch) resolution as a bitmap image consisting of 64,512 pixels (224 horizontal and 288 vertical). Each pixel has 8-bit-depth quantization level (gray-level images). The fingerprint digital image array acquisition process is completed by synchronous serial transfer to the application processor (DSP).

2.3 Feature Extraction Algorithms

To extract the features, a set of digital signal processing algorithms is applied to the captured utterance.

2.3.1 Voiceprint Hard Features

The following formulae have been applied to measure the voiceprint hard features [23]:

- Root mean square (RMS):

$$\mathrm{RMS}_j = \sqrt{\frac{1}{N}\sum_{m=0}^{N-1} s_j{}^2(m)} \tag{1}$$

- Zero-crossing rate (ZCR):

$$\mathrm{ZCR}_j = \sum_{m=0}^{N-1} 0.5\left|\mathrm{sign}\left(s_j(m)\right) - \mathrm{sign}(s_j(m-1)\right| \tag{2}$$

- Autocorrelation (AC):

$$\mathrm{AC}_j = \sum_{i=1}^{N}\sum_{j=1}^{N+1-i} s_j(j)\, s_j(i+j-1) \tag{3}$$

- Cepstral linear prediction coefficients (CLPC):

$$\mathrm{CLPC}_j = a_m + \sum_{k=1}^{m-1}\left(\frac{k}{m}\right) c_k a_{m-k} \tag{4}$$

with
$c_0 = r(0)$: first autocorrelation coefficient
a_m: prediction coefficients
$s_j(n) = w_N(n)s$: j-th windowed part of the uttered stream s
$w_N(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right)$: N-length weighting Hamming window

The above are short-time measurements executed by multiplying the N samples-wide weighting Hamming window $w(n)$ by the uttered speech stream $s(n)$.

RMS is the root mean square of the windowed speech segment. Such measurement helps to identify phonetic unit end-points.

ZCR is the time-domain measurement of the dominant frequency information. It is used to determine whether or not the current processed uttered speech segment is voiced or unvoiced and also to find the frequency (band) with the major energy concentration.

AC is the measurement of the speech-pitch frequency. It preserves information about pitch-frequency amplitude while ignoring phase. Phase is unimportant for the purpose of speech identification.

CLPC is the LPC-Cepstral feature vector that models the vocal tract.

2.3.2 Voiceprint Soft Features

The following soft features are extracted from speech:

- Speed.
- Stress.

Speed is measured as the total duration of the speech utterance.

Stress is measured as the ratio between the peak amplitude of the stressed vowel and the average amplitude of the whole utterance.

Both these voiceprint features are related to the way the person is used to speaking a requested word.

2.3.3 Fingerprint Hard Features

Several preprocessing steps are executed on the captured grayscale fingerprint image from the stage of bitmap to its minutiae-based representation:

- Orientation field and region of interest.
  - Normalization of the captured fingerprint image.
  - Image segmentation in blocks.
  - x and y gradient computation for each pixel in each block.
  - Local orientation for each pixel.
  - Orientation field correction.
- Positioning (delta and core localization).
- Ridging.
- Ridge thinning.

Fingerprint hard features are the minutiae. The criteria used for minutiae extraction from the thinned ridge image are the following:

- If a ridge pixel has two or more 8-nearest, then it is a bifurcation.
- If a ridge pixel has only one 8-nearest, then it is a termination.

As the crest-valley image is two levels encoded (1-0), the mapping algorithm is

$$\begin{aligned} &\text{if } \sum_{i=0}^{7} p_i > 2 \\ &\qquad \text{the pixel is B} \\ &\text{else} \\ &\qquad \text{the pixel is T} \end{aligned} \tag{5}$$

where p_i are the 8-nearest pixels of the pixel to be classified as bifurcation B or termination T.

The most critical step in this procedure is to avoid computing false minutiae caused by noise in the scanned fingerprint image.

To overcome this problem, a backtracking control is executed on each feature pattern before it is validated, to check that each of the three branches of the bifurcation is significantly long:

- Starting from the bifurcation all the three paths are checked, stopping if more than k pixels or a termination pixel is detected.
- If the stop is due to the termination detection, then the bifurcation and the terminations are invalidated.

The scanned fingerprint image is transformed into a set minutiae encoded by its x, y coordinates and the direction of the ridge corresponding at that position.

2.3.4 Fingerprint Soft Features

Two fingerprint soft features are extracted from fingerprint captured image:

- Total area.
- Mean intensity.

Both these two fingerprint features are related to the way the person approaches contact with the fingerprint sensor.

The total area is measured as the ratio between the total pixels available on the fingerprint scanning device and the total pixels of the captured fingerprint image that have a value higher than an estimated peak noise level. The peak noise level is estimated at calibration time and it is applied at run time (enrolling and identification).

The mean intensity is measured as average intensity of all the pixels with a value higher than a threshold level. The threshold level is 10 % higher than the peak noise level.

2.4 Hard Computing Pattern Matching

The captured fingerprint and voiceprint hard features have to be matched with enrolled features (templates).

2.4.1 Voiceprint Hard Features

Two methods are applied to score the person's identity. One is based on measuring Mahalanobis distance, and the other on measuring the distance of dynamic time warping—*k*-nearest neighbor (DTW-KNN).

The Mahalanobis distance measurement is

$$D_i(x) = (x - \bar{x})^T W^{-1} (x - \bar{x}) \tag{6}$$

where W is the covariance array computed using the average and the standard deviation features of the utterance. The input pattern x is processed with reference to the utterance-averaged feature vector $\bar{x}$ that represents the person to be identified. The distance $D_i(x)$ is a score for the authorized user.

The DTW-KNN (dynamic time warping—k-nearest neighbor) method combines the dynamic-time-warping measurement with the *k*-nearest neighbor decision algorithm. The DTW clusters similar

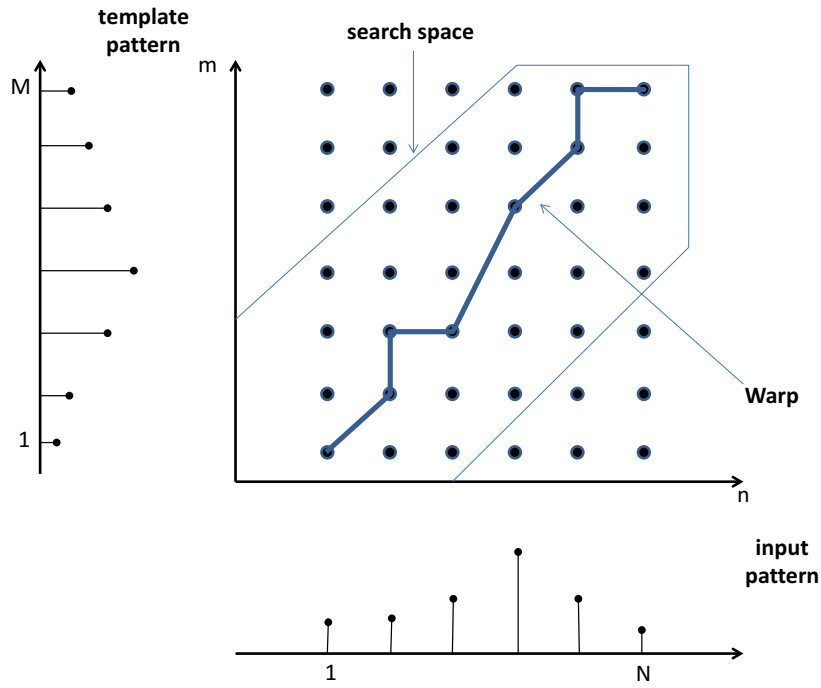

Fig. 1 DTW boundaries for voiceprint matching

elements that refer to a feature into classes (Fig. 1). The cost function is computed using Euclidean distance, with a granularity of one frame. The KNN algorithm is then applied to select k minimal distance matching and to choose the most recurring person in k minimal distance matches. This results in lower false-positive and false-negative rates during identification, compared to the original DTW algorithm.

2.4.2 Fingerprint Hard Features

Fingerprint hard feature pattern matching is based on minutiae. Pattern matching consists of a procedure that first tries to align the template pattern and the input pattern, and then computes an overlapping score (Fig. 2). The score is a measurement of the authenticity of the person who enrolled the input.

2.5 Soft Computing Pattern Matching

The captured fingerprint and voiceprint soft features have to be matched with enrolled features (templates). This is done using the artificial neural network (ANN) soft computing paradigm.

2.5.1 Artificial Neural Network (ANN)

The ANN is a signal processing and pattern classification paradigm inspired by the structure and functions of biological neural networks. Information signals that flow through the ANN modify the connections, enabling the learning process.

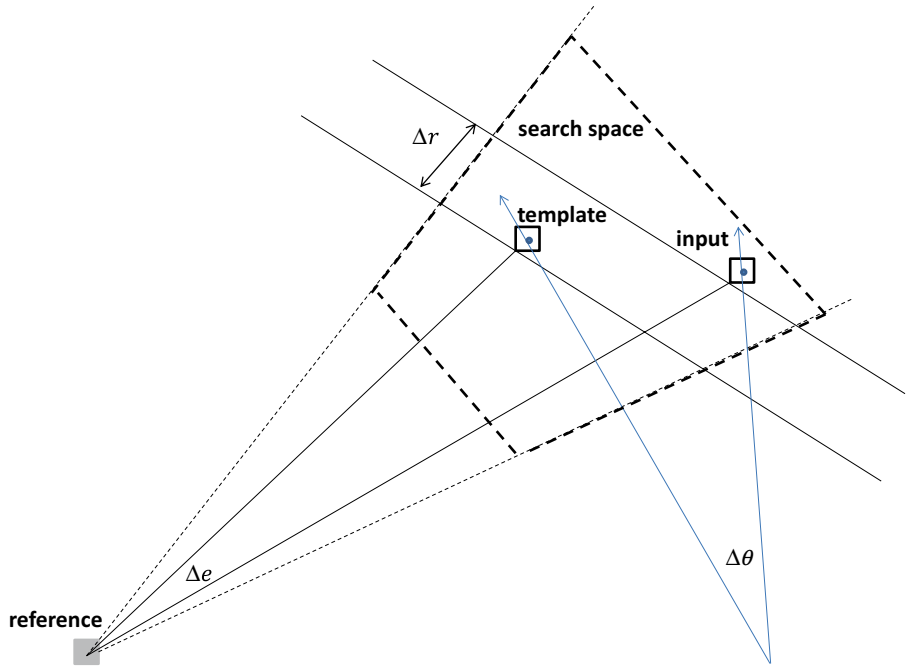

Fig. 2 Boundaries for fingerprint matching

The ANN applied to classify the soft biometric features is a feed-forward back-propagation neural network (FFBP-ANN) paradigm because of the behavioral nature of this kind of biometric data (Fig. 3a). It is a three-layered parallel processing network that processes the input data patterns at the lower layer, matches the data at the middle layer, and organizes the results at the upper layer. Its input nodes are fully connected to all the nodes in the hidden layer, and the hidden layer is fully connected to the output nodes. In this network the feed-forward action consists in that the i node at $l+1$th layer receives signals from the j node in the lth layer conditioned via weight w_{ij}^{l}. The activation of the node i at the $l+1$th layer is given by

$$f\left(\sum_{j=1}^{M} w_{ij}^{l} x_{j}^{l}(k) - \vartheta_{i}\right) \tag{7}$$

This is the activity of the ith node at the layer $l+1$ for the kth input x_j processed by the ANN, assuming that M nodes are in the lth layer.

Input and output layers have a linear activation function that controls the connection. A nonlinear (sigmoid) activation function

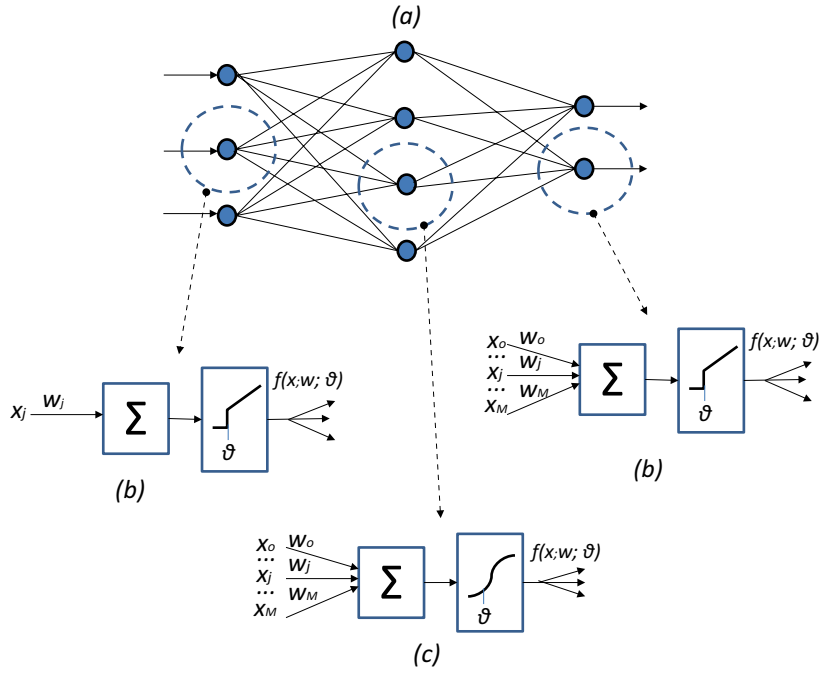

Fig. 3 (**a**) FFBP-ANN for voiceprint and fingerprint soft feature classification, (**b**) linear activation function for input and output layers, and (**c**) sigmoid activation function for hidden layer

(Fig. 3b) connects hidden-layer nodes to output-layer nodes according to the following formulae:

$$s_i = \frac{1}{1 + e^{-\beta I_i}}$$
$$I_i = \sum_j w_{ij} x_j - \vartheta_i \tag{8}$$

where ϑ is the activation threshold of the ith node and β is a constant that controls the slope of the semi-linear region of the sigmoid. When β is very small, the sigmoid approximates the linear activation function (Fig. 3c), so the same function can be applied to input, hidden, and output layers by choosing appropriate values for β constant.

2.6 Soft Computing Fusion and Decision

The fusion and decision logic is based on the fuzzy logic inferential paradigm. A fuzzy logic engine processes the crisp inputs (scores and features), applies to them a set of inferential rules, and makes the fuzzy decision. This is done by the fuzzy logic engine.

2.6.1 Fuzzy Logic Engine

The fuzzy logic engine (FLE) used to implement the data and decision fusion and to infer about the final decision is a zero-order Sugeno-type (Fig 4a). It fuzzyfies the inputs (soft features and scores)

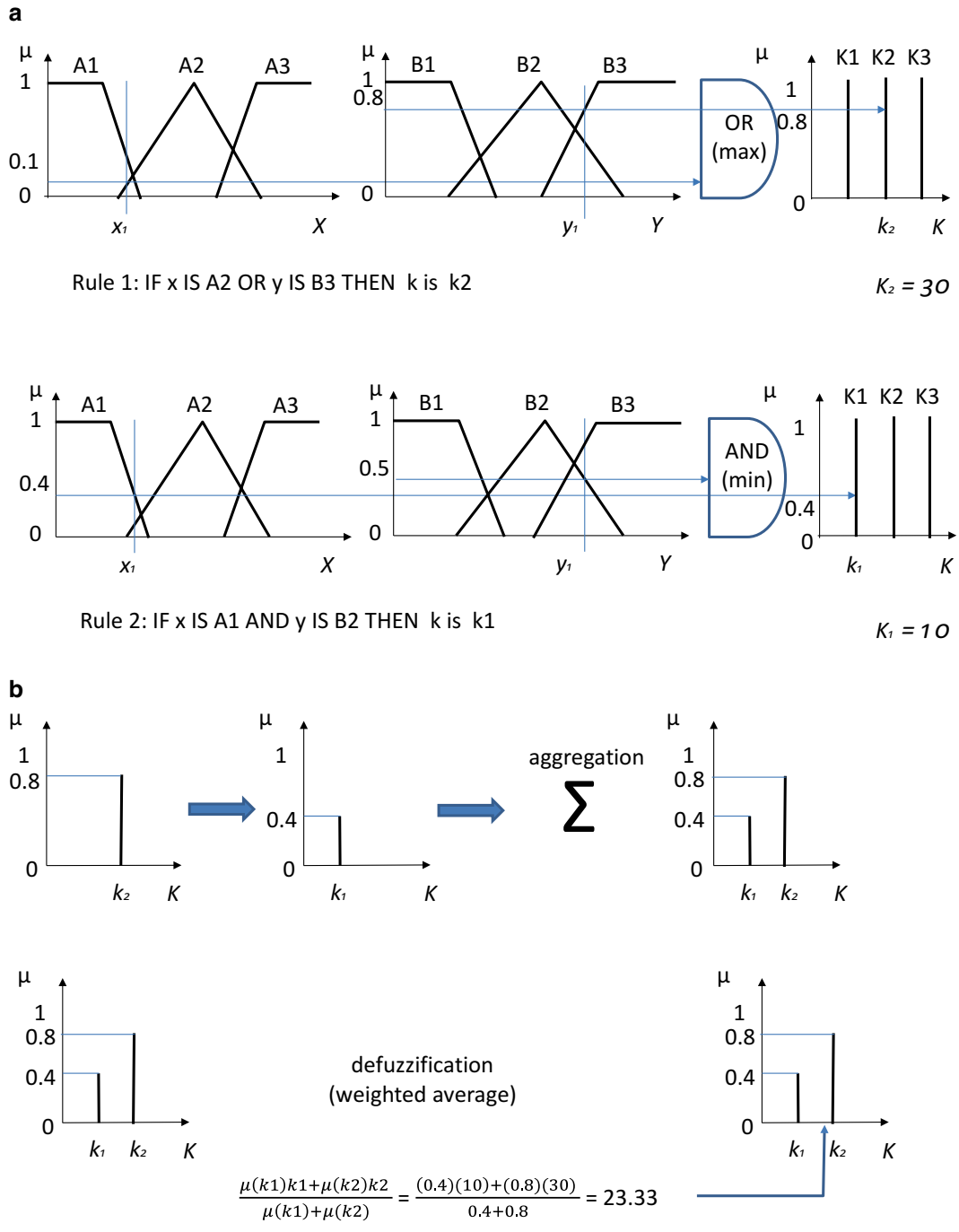

Fig. 4 (**a**) Zero-order Sugeno-type fuzzy logic inference engine. (**b**) Aggregation and defuzzification

using trapezoidal and triangular membership functions, and applies a set of inferential rules of the form:

IF *x IS A* AND *y IS B* THEN *k is K*
IF *x IS A* OR *y IS B* THEN *k is K*
IF *x IS A* THEN *k IS K*

The antecedent (input) part of the rule combines the fuzzyfied inputs by AND (minimum) fuzzy operator or by OR (maximum) fuzzy operator, or directly. The output of each rule is constant (zero order), and then the consequents are represented by singleton membership functions.

To defuzzify the output, so a crisp value is available, the weighted average (WA) method is applied (Fig 4b) (*see* **Note 3**):

$$\mathrm{WA} = \frac{\sum \mu(x)x}{\sum \mu(x)} \tag{9}$$

where $\mu(x)$ is the degree to which the inputs x belong to the appropriate fuzzy sets.

2.7 System Modeling and Simulation Environments

Matlab environment and the DSP toolbox are used to code the signal processing algorithms, to run them in simulation mode, and then to export the source code as ANSI-C. The Data Acquisition (DAQ) toolbox is used to collect data from the fingerprint sensor and from the microphone.

2.7.1 ANN

To model and simulate the FFBP-ANN the Matlab environment plus the DSP toolbox is used to code the FFBP-ANN, to run it in simulation mode, and then to export the source code as ANSI-C. The Data Acquisition (DAQ) toolbox is used to collect data from the fingerprint sensor and from the microphone.

2.7.2 FLE

The FLE is a Sugeno-type fuzzy logic inferential paradigm. To model and simulate it the Matlab environment plus the Fuzzy Logic toolbox is used to code the FLE, to run it in simulation mode, and then to export the source code as ANSI-C. The Data Acquisition (DAQ) toolbox is used to collect data from the fingerprint sensor and from the microphone.

2.8 Digital Signal Processor

To implement the embedded system, a system-on-chip (SoC) processor is used. This is an application-specific processor (ASP) optimized for signal processing, with on-chip memory and peripherals.

A 32-bit floating-point DSP, architecturally optimized to process data of different bit format (8-bit, 16-bit, 32-bit), has been chosen. Due to its computing architecture peculiarity, it is able to process efficiently the 8-bit data of the image pixels and the 16-bit data of the voice samples. The 32-bit floating-point data format ensures the required computing precision both for the image and the speech processing.

3 Methods

The biometric authentication system (Fig. 5) consists of three processing layers, the feature-extraction layer, the matching layer, and the fuzzy logic-based decision layer. The feature-extraction layer uses signal processing-based algorithms for hard and soft feature extraction. The matching layer uses the hard computing (DSP) to identify the hard features and the soft computing (ANN) to identify the soft features. The decision layer uses the soft computing (FLE) to fuse the identification scores with the soft features.

3.1 Feature Extraction Layer

The feature extraction layer implements the biometric information capture (voiceprint and fingerprint) and the signal processing algorithm (hard computing) methods for feature extraction.

3.1.1 Voiceprint and Fingerprint Capture

The voiceprint is captured by a voice-type microphone, which is conditioned (linearized, amplified, and filtered), in the analog domain and then 16-kHz sampled, and 16-bit quantized. The utterance is optimally end-pointed [21] and segmented. The Hamming window is applied to extract 10-ms frames from the speech-data stream, using a 50 % overlap between adjacent frames to avoid data loss at feature-extraction time.

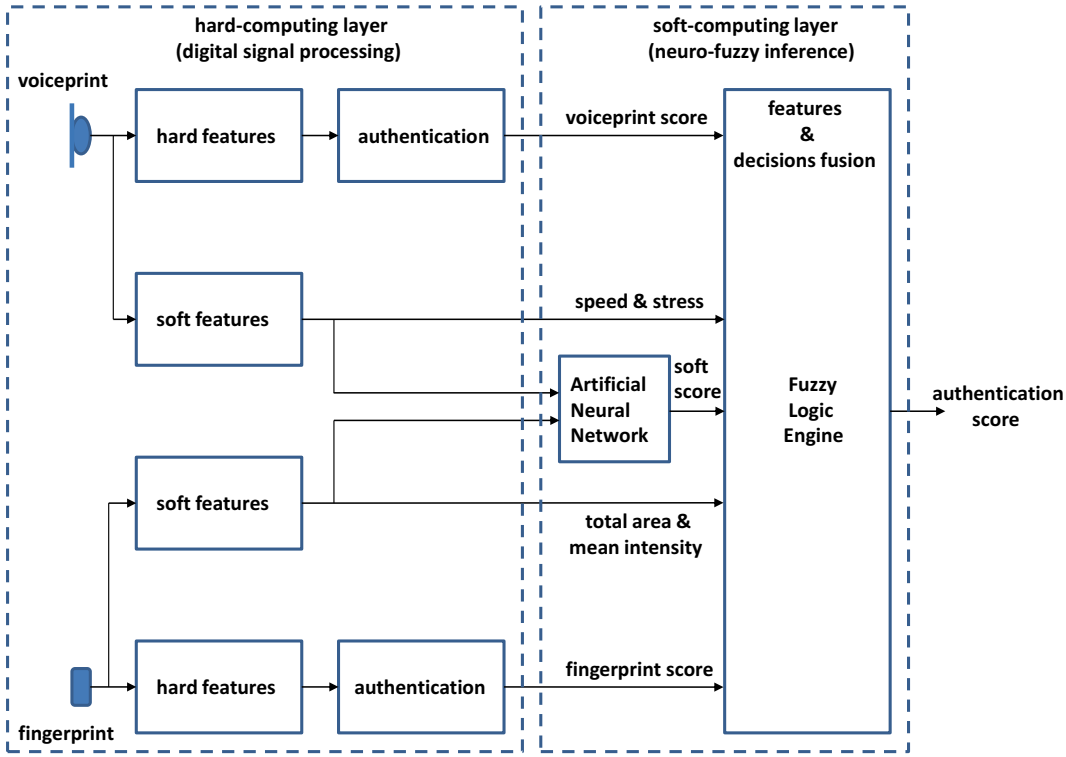

Fig. 5 System architecture

The fingerprint image is captured by a 512 dot-per-inch (dpi) solid-state fingerprint sensor. The image is available at the sensor output as 64,512 pixels 8-bit quantized array (224 rows and 288 columns) image.

3.1.2 Feature Extraction

Features (hard and soft) are extracted from the captured voiceprint and fingerprint applying the hard computing algorithm. Data are assembled in data structures useful to be processed at the pattern matching layer.

Voiceprint hard features are structured as time-sequenced vectors of data:

RMS(N)

ZCR(N)

AC(N)

CLPC(N)

where N is the vector length and each vector element is the feature measurement at the window application time.

Fingerprint hard features are structured sequences of minutiae as:

MINUTAE(M)

where M is the vector length and each vector element is the minutiae data.

From voiceprint the two soft features are measured as:

SPEED

STRESS

The two features are scalar data (one measurement executed on the whole captured voiceprint data stream).

From the fingerprint two soft features are measured as:

TOTAL_AREA

MEAN_INTENSITY

The two features are scalar data (one measurement executed on the whole captured fingerprint data array).

3.2 Matching Layer

The matching layer implements the hard and soft computing scoring systems. Each of these applies the appropriate algorithms for scoring the input data referred to a set of template data belonging to the persons to be identified. This layer runs in two different modes, enrolling and identifying.

The enrolling mode is active when a new person is to be added to the set of persons that are to be accepted. The identifying mode is active when a person (allowed or not allowed) is accessing the system furnishing its voiceprint and fingerprint. Both enrolling and identification modes run the same pattern matching algorithms, the first to store the reference templates, and the second to execute the scoring.

3.2.1 Enrollment Mode

Voiceprint enrollment mode consists in storing as multiple templates the features measured each time the allowed person utters at the microphone a specific (requested) vocal sequence. Fingerprint enrollment mode consists in storing as multiple templates the features measured each time the allowed person puts a finger (requested) on the sensor.

3.2.2 Identifying Mode

The voiceprint and fingerprint identifying mode runs the pattern matching algorithms to score the input hard features related to the stored templates. For each input (voiceprint and/or fingerprint) a score is available at the matching layer output.

3.2.3 Soft Biometric Training and Scoring

The soft biometric scoring is executed running the FFBP-ANN. To do this the FFBP-ANN inputs the soft biometric features and outputs the score according to how it learned about the soft feature belonging to the authorized person. A training phase is requested to embed the knowledge in the FFBP-ANN nodes (neurons). This is run each time the soft features at the FFBP-ANN inputs belong to an authorized person, explicitly during the enrollment (training mode), and implicitly each time the person is identified as authorized (evolving mode). Learning is executed running the error back-propagation algorithm.

Back Propagating Networks (BPNs) are trained according to a generalized least mean-square (LMS) algorithm:

$$w_j(k+1) = w_j(k) + \eta\left(x^t(k) - x(k)\right) f_j(k) \tag{10}$$

The weights $w_j(k)$ are modified by the kth input activity pattern $f_j(k)$, so that a new updated weight $w_j(k+1)$ is available at $k+1$th input time. The modification is proportional to the difference between the $x^t(k)$ target response and the current response $x(k)$. The constant η controls the learning rate and is in the range $0 < \eta < 1$.

During the learning activity the difference between the target activity and the current activity at the output layer is a learning error indicator. To measure it, the total error E is measured as

$$E = \sum_{k=1}^{K} \frac{1}{2} \sum_{i=1}^{N} (x_i^t(k) - x_i^L(k))^2 \tag{11}$$

This is the total error measured at the N nodes of the output layer L after K patterns of the training set have been applied. If E is minimum, then weights at the end of the training are the best for the *L-layer* BPN.

Applying the best trained weights set to the three-layer FFBP ($L=3$) enables this ANN to perform the scoring of the biometric features. The ANN executes also the data fusion among the soft biometric data from voiceprint and fingerprint.

3.3 Decision Layer

The decision layer implements the data fusion and the decision fusion information from the matching layer. This is executed by the FLE that evaluates four kinds of data input:

- Voiceprint score.
- Fingerprint score.
- Soft-biometric measurements.
- Soft-biometric score.

Prior to a run, the FLE needs to be tuned for processing the data inputs and producing the decision. Membership functions and rules set have to be designed using the modeling and simulation toolbox.

Input data are fuzzyfied using optimally tuned membership functions (Figs. 6, 7, and 8a). Singleton membership functions are applied for rule consequents (Fig. 8b).

The rules are tuned combining the fuzzyfied inputs as follows (only the most significant are reported):

1. *IF voiceprint_score IS high AND fingerprint_score IS high AND soft_score IS high THEN authentication IS very high.*
2. *IF voiceprint_score IS medium AND fingerprint_score IS medium AND soft_score IS high THEN authentication IS high.*
3. *IF voiceprint_score IS medium AND speech_speed IS high AND soft_score IS medium THEN authentication IS high.*

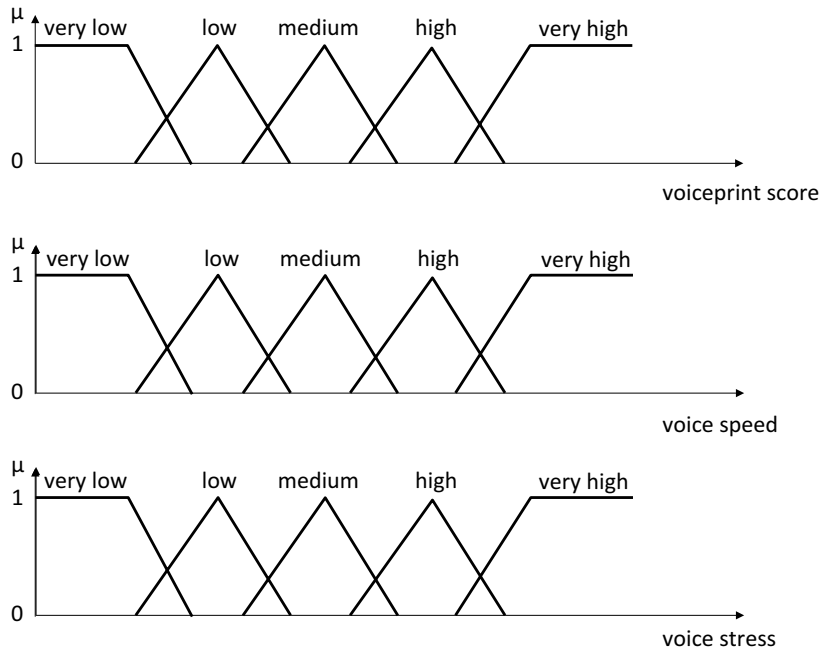

Fig. 6 Membership functions to fuzzify inputs (voiceprint)

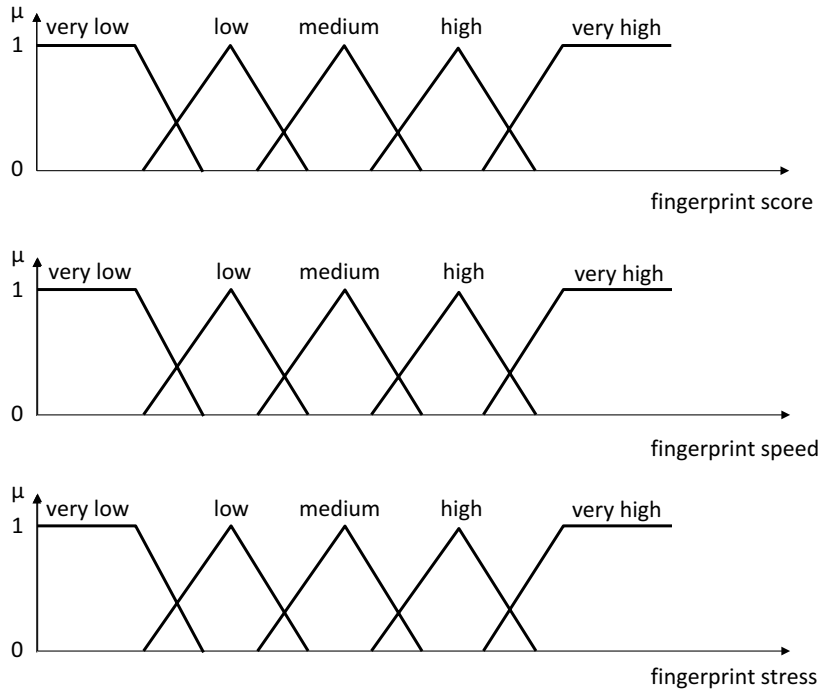

Fig. 7 Membership functions to fuzzify inputs (fingerprint)

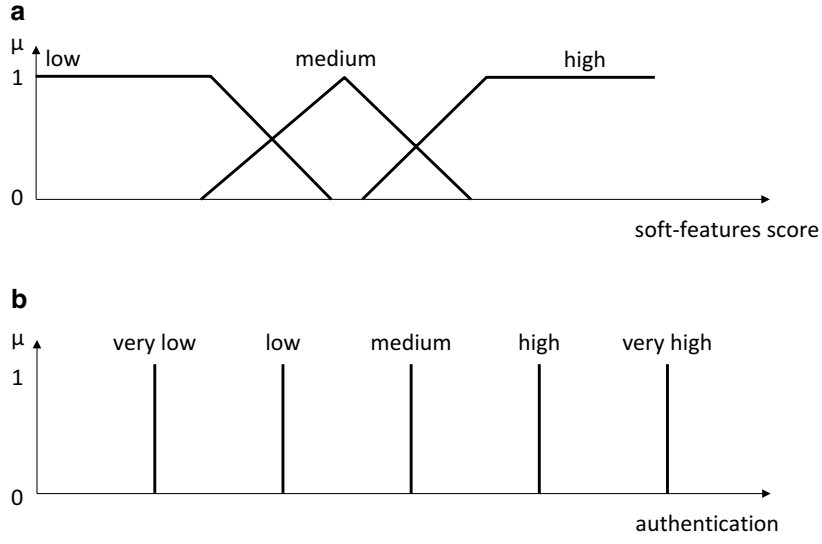

Fig. 8 (**a**) Membership functions to fuzzify inputs (soft-features score) and (**b**) to defuzzify outputs (authentication)

4. *IF voiceprint_score IS low AND speech_speed IS high AND speech_stress IS high THEN authentication IS average.*
5. *IF fingerprint_score IS medium AND fingerprint_total_area IS high THEN authentication IS high.*
6. *IF finger_print_score IS low AND fingerprint_total_area IS high AND fingerprint_mean_intensity IS high THEN authentication IS average.*
7. *IF voiceprint_score IS medium and fingerprint_score IS medium AND soft_score IS low THEN authentication IS low.*
8. *IF voiceprint_score IS high and fingerprint_score IS low AND soft_score IS low AND fingerprint_total_area_match IS low AND fingerprint_mean_intensity_match IS low AND speech_speed_match IS low AND speech_stress_match IS low THEN authentication IS very low.*
9. *IF voiceprint_score IS low and fingerprint_score IS high AND soft_score IS low AND fingerprint_total_area_match IS low AND fingerprint_mean_intensity_match IS low AND speech_speed_match IS low AND speech_stress_match IS low THEN authentication IS very low.*

Rules were derived from feature distribution. Each rule was manually tuned using a fuzzy logic rule editor, the simulator, and the knowledge of an expert:

- Rule 1 is a reinforcement of voiceprint and fingerprint matchers.
- Rule 2 combines a voiceprint matcher and a fingerprint matcher when both scores are too close to the decision threshold.
- Rules 3, 4, 5, and 6 act as recovery rules when the voiceprint or fingerprint matchers generate a false rejection.
- Rules 7, 8, and 9 protect against false acceptance.

For fine-tuning, many other rules can be generated to take additional soft-biometric measurements into account. Using more rules leads monotonically toward greater reliability in the authentication process.

Trapezoidal and triangular membership functions are used to process inputs. The inference rule set is then applied. The result of all the rules is evaluated using the WA method, so the crisp output value can be computed. Singleton membership functions were used to defuzzify the final decision.

3.4 Performance Evaluation

Performance evaluation aims to measure the reliability of the implemented biometric identification method. Voiceprint and fingerprint authentication were first implemented and tested separately, and then jointly, and finally combined through the fuzzy logic inference engine and the artificial neural network applied to soft-biometric features. An "all-against-all" test strategy was applied to obtain match and mismatch scores.

Evaluation of joint voiceprint and fingerprint authentication consists of taking as good the better of the two matches (OR). Single-user authentication was performed, so this test had minimal system requirements. The following results were produced:

- Voiceprint alone: 90.5 % correctly accepted.
- Fingerprint alone: 85.7 % correctly accepted.
- Voiceprint OR fingerprint: 92.3 % correctly accepted.
- Fuzzy logic decision fusion of voiceprint, fingerprint, and artificial neural network evaluated soft features: 95.8 % correctly accepted.

The OR test confirmed that multi-biometrics can improve performance compared to single-biometric authentication. System performance can be significantly improved, while keeping complexity to a minimum, using fuzzy logic as decision fusion and reinforcing it with an artificial neural network applied to soft-biometric features.

4 Notes

1. A dual microphone or an array of microphones needs to be used when the biometric application is targeted to outdoor applications and the person to be identified is close to other persons. Beam forming can be implemented for noise reduction purpose.
2. The capacitive fingerprint sensors are implemented as a two-dimensional or a mono-dimensional (strip) scanner. The performance of the biometric system is not sensitive to this form factor. The first is less computationally intensive and more intuitive, but it is not optimal for system dimension reduction. The second is computationally intensive because it implies that the two-dimensional image has to be built by software and it is less intuitive, but it is optimal for system dimension reduction.
3. Weighted Average (WA) defuzzification method is a derivation of the Centroid (Center of Gravity) method that fits well the singleton membership function.

References

1. Jain AK, Pankanti S, Prabhakar et al (2004) Biometrics: a grand challenge. Proc 17th international conference on pattern recognition (ICPR), vol 2. pp 935–942
2. Anil K, Jain AK (2012) Biometric recognition: an overview. In: Mordini E, Tzovaras D (eds) Second generation biometrics: the ethical, legal and social context. The International Library of Ethics, Law and Technology, vol 11. Springer, pp 49–79
3. Baird SL (2002) Biometrics: security technology. The technology teacher 61(5):1, pp 8–22

4. Sujithra M, Padmavathi G (2012) Next generation biometric security system: an approach for mobile device security, Proc second international conference on computational science, engineering and information technology. ACM, New York, NY, USA, pp 377–381
5. Corcoran P, Cucos A (2005) Techniques for securing multimedia content in consumer electronic appliances using biometric signatures. IEEE Trans Consumer Elect 51:545–551
6. Jain AK, Ross A, Pankanti S (2006) Biometrics: a tool for information security. IEEE Trans Inf Forensics Security 1(2):125–143
7. Jain AK, Nandakumar K, Lu X et al. (2004) Integrating faces, fingerprint, and soft biometric traits for user recognition. Proc biometric authentication workshop, LNCS 3087, Prague, pp 259–269
8. Hong A, Jain S, Pankanti S (1999) Can multibiometrics improve performance? Proc AutoID`99, Summit, NJ, pp 59–64
9. Anil J, Lin H, Sharath P (2000) Biometric identification. Commun ACM 43(2):90–98
10. Cappelli R, Maio D, Maltoni D et.al (2006) Performance evaluation of fingerprint verification systems. IEEE Trans Pattern Anal Mach Intell 28:3–18
11. Bishop C (1995) Neural networks for pattern recognition. Oxford University Press, Oxford
12. Ciota Z (2001) Improvement of speech processing using fuzzy logic approach. Proc of IFSA World congress and 20th NAFIPS Int Conf, 2
13. Bosteels RTK, Kerre EE (2007) Fuzzy audio similarity measures based on spectrum histogram and fluctuation patterns. Proc Int Conf Multimedia and Ubiquitous Engineering 2007, Seoul, Korea, 27–28 April
14. Malcangi M (2002) Soft-computing approach to fit a speech recognition system on a single-chip. Proc 2002 international workshop system-on-chip for real-time applications, Banff, Canada, 6–7 July
15. Malcangi M (2004) Improving speech endpoint detection using fuzzy logic-based methodologies. Proc thirteenth Turkish symposium on artificial intelligence and neural networks, Izmir, Turkey, pp 10–11
16. Wahab A, Ng GS, Dickiyanto R (2005) Speaker authentication system using soft computing approaches. Neurocomputing 68:13–17
17. Kar B, Kartik B, Dutta PK (2006) Speech and face biometric for person authentication. Proc IEEE international conference on industrial technology, India, pp 391–396
18. Jang-Hee Y, Jong-Gook K et al (2007) Design of embedded multimodal biometric systems. Proc of the int conf on signal image technologies and internet based systems, SITIS 2007:1058–1062
19. Jain AK, Dass SC, Nandakumar K (2004) Soft biometric traits for personal recognition systems. Proc international conf on biometric authentication, LNCS, vol 3072. Hong Kong, pp 731–738
20. Pak-Sum Hui H, Meng HM, Mak M (2007) Adaptive weight estimation in multi-biometric verification using fuzzy logic decision fusion. Proc IEEE international conference on acoustic, speech, and signal processing ICASSP'07, vol 1. pp 501–504
21. Runkler TA (1997) Selection of appropriate defuzzification methods using application specific properties. IEEE Trans Fuzzy Sys 5: 72–79
22. Abiyev RH, Altunkaya K (2007) "Neural network based biometric personal identification", Frontiers in the Convergence of Bioscience and Information Technologies. pp 682–687
23. O'Shaugnessy D (1987) Speech communication—human and machine. Addison-Wesley, Reading, MA