Dempster-Shafer-Based Fusion of Multi-Modal Biometrics for Supporting Identity Verification Effectively and Efficiently

Alfredo Cuzzocrea*
iDEA Lab, University of Calabria, Rende, Italy
and LORIA, Nancy, France
alfredo.cuzzocrea@unical.it

Enzo Mumolo
Department of Engineering
University of Trieste, Trieste, Italy
mumolo@units.it

***Abstract*—The purpose of this paper is to describe a novel fusion algorithm for multi-modal biometric identification. In this paper we describe the fusion of fingerprints and voice. This combination of biometrics is rarely used in verification systems although this biometric pair is simple to use and not too invasive. A framework for the combination of several data fusion algorithms is described. In this paper we use only two types of data fusion techniques, namely weighted sum and fuzzy system. Two independent identity decisions can be thus obtained, and from them two beliefs that the identity is verified can be derived. The two beliefs are combined using Dempster-Shafer's approach to obtain the final decision. The results are reported by ROC curves.**

***Index Terms*—Biometrics, Multi-Modela Biometrics, Identity Verification.**

I. Introduction

In this paper we address the problem of identity verification using bi-modal biometries namely voice and fingerprint. Biometrics are information of biological origin that belongs only to a person and is not shared by others. We could say that practically almost all the biological characteristics of a person that can be measured by a currently available technique characterize only that person. DNA, iris, face, fingerprint, signature or the voice are some of the possible characteristics of the identity of each person. When acquired and compared, they can be used to distinguish people themselves. This distinction can be used to both identify and verify the identity of a person. The difference between identification and verification lies in the fact that in the first case the system recognizes who that biometrics belongs to (among a set of known biometries) and in the second the system ascertains or not if a person is who he/she claims to be. In any case, the biometric characteristics previously acquired and stored are compared with that produced in that precise moment.

Multi-modal biometrics integrates several biometric information using data fusion techniques. There are two main reasons for using multi-modal biometrics. The first is that the uni-modal biometries can be disguised in some way; hence the use of multi-modal biometrics should make camouflage more difficult. The second is that no single biometrics can guarantee perfect verification of a person's identity. Using at least two different biometrics should make it possible to compensate for the limitations of individual biometries. It is expected that one good integration of different biometries improves overall results.

The main objective of an identity verification system is to control the access to resources that cannot be shared with many people. For this reason such systems should minimize the rate of false positive which is the percentage of times the system accepts impostors. A multi-biometric system has one or more inputs linked to different biometrics, and a binary answer: to accept or reject the identity. The different modes (for example face image, iris image, fingerprint or voice) that represent the same identity from different perspectives may be integrated at four different fusion levels, namely at the sensor, at the feature, at the score and at the decision levels as described in [1], [2], [3], [4].

Extensively research has been done on face recognition, which is a very natural biometry for human beings. However, face images recognition performance can be limited by many factors, namely lighting conditions, poor image quality, image size and face angle. In [5] the performance of some face recognition algorithms based on Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are compared. In [6] a survey of methods of face recognition is reported. To overcome the limits of face verification, face biometry is commonly used together with other biometries. For example in [7] a bi-modal verification system is described using face and fingerprint. A comparison of different fusion approaches, namely sensor, feature, score and decision is carried out showing that fusion at the score level leads to the best results. In [8] a verification system based on face and fingerprint is described. The authors show that verification results using sum-based fusion are better than face or fingerprint alone.

Another very natural biometry for human beings is voice. Also speaker verification by voice has been studied extensively. However, also speaker verification can be limited by many factors. For example it strongly depends on the psyco-

*This research has been made in the context of the Excellence Chair in Computer Engineering – Big Data Management and Analytics at LORIA, Nancy, France physical condition of the speaker and on the acoustic noise in the environment. It can be morover easily subjected to voice mimicry attacks. Most speaker verification systems are based on a Universal Background Model (UBM) which is a text-independent, gender-specific pool of different speakers. Each speaker is represented by acoustic feature vectors modeled by Gaussian Mixture Model (leading to the popular GMM-UBM model) as described in [9] and, more recently, by *i-vectors* as reported in [10] or *x-vectors* as described in [11].

Many multi-modal identity verification systems has been developed using the speech biometry. For example in [12] three types of deep neural networks which integrate face and voice are described. Identity verification experiments conducted by the authors with the VoxCeleb dataset show that better Equal Error Rate (EER) results with respect to the state of the art are obtained. In [13] four fusion strategies are tested with face and voice. Performance results obtained by the authors using their own dataset show that better EER values are obtained with feature level fusion. Moreover, Zhang *et al* in [14] describe an application for smartphone unlocking. The fusion is performed at the score level.

Applications of biometric identity verification are rapidly evolving, from the development of electronic passports, to the access of health services, to banking, to electronic voting or to the control of physical or logical devices to guarantee the security of access. The devices that are becoming increasingly important are personal computers and smartphones, which open to a vast amount of important personal data. For example Memon describes in [15] a multi-biometric access control of smartphones using the combination of three biometrics, namely fingerprints, face and voice. The fusion of the three biometries is obtained with Support Vector Machine (SVM) which must be initially trained.

The biometries considered in our paper are voice and fingerprint which is a bi-modal combination that has not been studied very thoroughly. One of the few papers dealing with fingerprint and voiceprint based identity verification is [16] where the authors describe the combination of the two biometries by fusing the two biometries at the feature level. In [16], voice is modeled by means of GMM and fingerprint are represented with minutiae, which is similar to our case.

One contribution of this paper is that a framework for the combination of several data fusion algorithms is proposed. In fact it is not possible to know a-priori what is the best suited data fusion algorithm, so we transform the output of data fusion algorithms to belief functions, to obtain several independent opinions on the identity verification which can be combined. Another contribution is that we propose a simple fingerprint matching algorithm based on dynamic programming.

II. Algorithm Overview

The goal of the verification algorithm is to evaluate the probability related to two events:

H_1 which means that the identity is verified

and

H_0 which means that the identity is not verified

We describe in the following the proposed framework for the fusion of decisions resulting from multiple data fusion algorithms among the biometric scores. The framework is illustrated in Fig. 1. In the Figure we consider only two types of biometries and K different types of data fusion algorithms as an example.

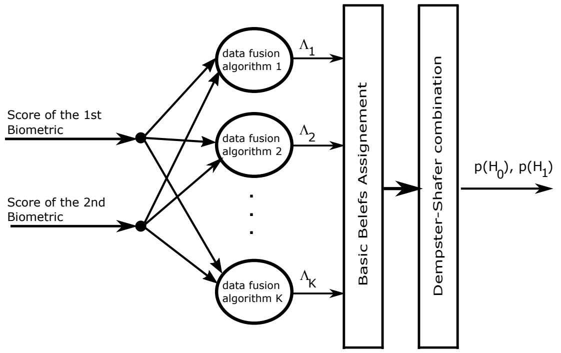
Score of the 1st Biometric
Score of the 2nd Biometric
data fusion algorithm 1
data fusion algorithm 2
data fusion algorithm K
Λ1
Λ2
ΛK
Basic Beliefs Assignement
Dempster-Shafer combination
p(H0), p(H1)

Fig. 1. Block diagram of the proposed decision fusion framework.

The Dempster-Shafer (DS) theory of evidence [17] represents uncertainty and lack of knowledge combining different measures of evidence. The DS approach considers a finite set of hypotheses, in our case $\Theta = \{H_0, H_1\}$. This assumption leads to the following set of $2^{|\Theta|}$ possibilities: $\{\emptyset$, $\{H_0\}$, $\{H_1\}$, $\{H_0, H_1\}\}$.

The Basic Belief Assignment (BBA) is a function $m(.)$ which associates to every subset $\mathcal{A}$ of the hypothesis set Θ a value in the range [0, 1]. $m(.)$ shall satisfy the following conditions: $\sum_{\mathcal{A} \subseteq \Theta} m(\mathcal{A}) = 1, \;\; m(\emptyset) = 0$. The belief function, $bel(.)$, associated with the Basic Belief Assignment $m(.)$, assigns a value in [0, 1] to every nonempty subset $\mathcal{B}$ of Θ. It is defined by

$$bel(\mathcal{B}) = \sum_{\mathcal{A} \subseteq \mathcal{B}} m(\mathcal{A}) \tag{1}$$

The belief function can be viewed as a generalisation of a probability function.

In our case the first operation to do is the conversion of the scores resulting from the data fusion systems (i.e. the $\Lambda_1, \ldots, \Lambda_K$ in the figure) into values, we call p_1, …, p_K, which lie between 0 and 1 and are the likelihood that the identity is verified or not. One simple way to perform this conversion is to measure the difference between the values of the Λ_i from the threshold θ related to each data fusion algorithm. The computation of the thresholds for the two data fusion algorithms we used in this paper is reported next. In summary, each data fusion algorithm becomes an Expert in DS fusion.

Under this assumption, the expert j provides the probability of the subset θ_i: $m_j(\{\theta_i\}) = p_i$.

The probability of the subset θ_i will be distributed to all the other subsets of Θ: for all the other subsets $\mathcal{C}$ of Θ, $m_j(\mathcal{C}) = \frac{1-p_i}{2^K-1}$.

For example, if K=3, each expert shall assign a belief to the subsets of Θ, $2^2 = 4$ sets. Thus, the j expert assigns $m_j(\{\theta_i\}) = p_j$ and $m_j(\mathcal{C}) = \frac{1-p_j}{3}$, $\mathcal{C} \subseteq \Theta$, $\mathcal{C} \neq \{\theta_j\}$.

Two BBAs $m_1(.)$ and $m_2(.)$ can be combined using the following formula [17]:

$$m(\mathcal{C}) = m_1 \bigoplus m_2 = \frac{\sum_{j,k,\mathcal{A}_j \cap \mathcal{B}_k = \mathcal{C}} m_1(\mathcal{A}_j) m_2(\mathcal{B}_k)}{1 - \sum_{j,k,\mathcal{A}_j \cap \mathcal{B}_k = \emptyset} m_1(A_j) m_2(B_k)} \tag{2}$$

where $\mathcal{A}_j$ and $\mathcal{B}_k$ are subsets of Θ. More than two BBAs can be combined in the same way.

The whole algorithm proposed in this paper is represented in Fig. 2, where two biometries and two data fusion algorithms are used. The users are divided into two groups, the authorized to access the resource and the non-authorized users, or impostors. Voice and fingerprints biometrics are acquired by the same person, and are represented with Gaussian Mixtures and with a list of minutiae respectively. The two representations of authorized persons are compared with the voice and fingerprints relating to the set of impostors in order to obtain scores, called Λ, with the following property: if the person is authorized to access the resource, his score will have a high value, otherwise he/she will have a low value. The determination of the scores will be described in the next sections.

In summary, the research performed so far indicate that generally the fusion at score level is preferable. However there is no indication on what is the fusion algorithm to use. Therefore, the two biometries are combined with two independent data fusion algorithms, namely weighted sum and fuzzy fusion algorithms, in order to obtain two different decisions at different likelihood level. A final decision is performed on the basis of Dempster-Shafer combination rule.

III. Voice Biometric

The GMM models of speech, described for example in [18], is a mixture of Gaussian distribution. A GMM model is denoted as

$$\lambda = \{w_i, \mu_i, \Sigma_i\},\ i = 1, \dots, M \tag{3}$$

where w_i are the mixture weights, μ_i the mean of each Gaussian and Σ_i the covariance of each Gaussian. Each speaker is represented by its own λ. In the following we briefly highlight the algorithm used in this paper. First, speech is split in short frames. Each frame is then parameterized in a feature vector x composed by Mel Frequency Cepstral Coefficients. The mixed density of the feature vector with a D dimension is a weighted linear combination of M uni-modal Gaussian probability densities, p_i, each parameterized by a vector of the means:

$$p(x|\lambda) = \sum_{i=1}^{M} w_i p_i(x) \tag{4}$$

where the density is reported in (5)

$$p_i(x) = \frac{1}{(2\pi)^{D/2} Det(\sum_i)^{1/2}} e^{\frac{1}{2}(x-\mu_i)(\Sigma_i)^{-1}(x-\mu_i)} \tag{5}$$

The weights of the mixtures, w_i must satisfy the condition $\sum_{i=1}^{M} w_i = 1$.

In the training phase, the goal is to estimate the GMM parameters to create a speaker model, which best approximates the vector distribution of the training characteristics. The most popular technique is Maximum Likelihood ML, the purpose of which is to find the parameters of the model that maximizes the likelihood of the GMM, given a series of Training vectors $X = \{x_1, \dots, x_T\}$. The likelihood of the GMM model is given by (6).

$$L_{GMM} = \log p(x|\lambda) \tag{6}$$

This function is not linear of λ and direct maximization is not possible. However the ML estimation of the parameters can be obtained iteratively using a special case of the EM (Expectation - Maximization) algorithm. The basic idea is, starting with an estimate of λ, to estimate a new model $\overline{\lambda}$ that allows to have $p(x|\overline{\lambda}) > p(x|\lambda)$. The new model becomes the starting model for the next iteration. The process is repeated a specified number of times. Two critical factors in training are the selection of the M order of the mixtures and the a priori initialization of the model parameters before the EM algorithm. The first problem can be solved experimentally, while in the second case it is noted that, from experimental evidence, elaborate initialization schemes are not necessary for the training of the speaker models. We choose the K-means algorithm, which allows clustering from the training data, minimizing the global mean distortion $D = E[d(x,z)]$. The idea is to divide the set of training vectors into L groups C_i, such that the following conditions are satisfied:

- The optimal quantizer is selected using a selection criterion for maximum proximity, formally $q(x) = z_i \iff d(x,z_i) \leq d(x,z_j) \forall j \neq i : 1 \leq j \leq$
- Each centroid z_i is chosen to minimize the average distortion in the corresponding c_i.

The algorithm can be described in the following steps:

1) Initialization: method for constructing an initial set of centroids. We start with a cokebook that uses the first $2d$ elements.
2) Classification: each training vector x is classified in a region C_i, choosing the nearest centriode z_i, that is $x \in C_i \iff d(x,z_i) \leq d(z,z_j) \forall j \neq i$
3) Centroid library update: calculated the new centroid for each region C_i from the training vectors of the region itself
4) End: if the decrease of the global distortion is below a certain predetermined threshold or the cycle has been

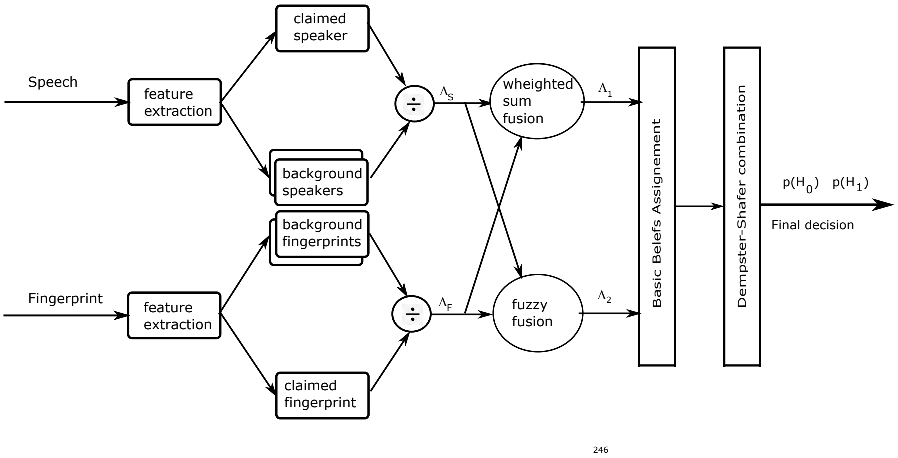

Fig. 2. Block diagram of the proposed bi-modal verification algorithm.

iterated for a certain number of times, then the procedure is concluded, otherwise you go back to step 2.

Using GMM modeling of voiceprints, we compute the GMM model of the Authorized person λ_A^S, and the average GMM model of Non-Authorize Persons, λ_{NA}^S, so we can compute the score

$$\Lambda_S = \frac{p(x|\lambda_A^S)}{p(x|\lambda_{NA}^S)} \tag{7}$$

where x is a segment of Speech and the S apex stands clearly for Speech. Equation (7) in logarithmic terms, becomes:

$$\overline{\Lambda_S} = \log \Lambda_S = \log P(x|\lambda_A^S) - \log P(x|\lambda_{NA}^S) \tag{8}$$

This term is the score used in the data fusion algorithms. This score will be very high if the x speech segment comes from the owner of the model λ_A^S, while it will be very low if the x speech segment is pronounced by an impostor.

IV. Fingerprint Biometric

We used a variant of the verification approach based on minutiae detection, which is reported for example in [19]. The algorithm is composed of the following points:

1) Acquisition of fingerprints In the system described in this paper, the acquisition of the fingerprints are made with n scan sensor shown in Fig. 3 The use of this sensor considerably simplifies the matching of the fingerprints because the physical structure of the sensor constraints the fingers to scan always at the same orientation.

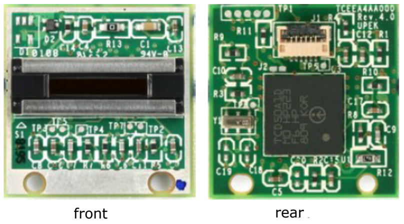

Fig. 3. Sweep-type fingerprint scanner Upek TCEEA4.

2) Determination of the direction of the ridges The fingerprint image is divided into blocks of $w \times w$ pixels (for example 16×16) pixels, so for each block centered in the point $P(i, j)$, a window of size $l \times w$ is defined (for example 32 $times16$) . Inside each block, the gradient $dx(i, j)$, $dy(i, j)$ is calculated for each pixel of coordinates (i, j), in turn used to calculate an estimate of the local orientation of each block. Extending the calculation to the whole image, we obtain a vector field where each vector is associated with a single block.
3) Determination of the frequency of the ridges.
In a generic subsection of the image, where there are no singularity points or minutiae, the ridges and valleys that make up the footprint can be modeled as a sine wave which propagates in a direction normal to the direction of the ridge itself. Each image blocks of $w \times w$ pixels is oriented according to the direction of the crest in the point obtained in the previous step. Ultimately the function returns a field (map) of frequency values point by point.
4) Creating the mask This operator constructs an image of the same size as the original footprint, but made up of only two values: it colors the areas of the image considered valid in white and the damaged ones in black. In the following steps, the operations on the impression will take into account the mask and will avoid working on the areas marked in black, thus saving calculation time and improving the accuracy of the results. To decide whether a pixel is valid or not, this function is based on the crest frequency map. The mask thus obtained is filtered to better define the edges and to fill any small holes formed by mistake in the critical central parts of the impression, for example in correspondence with the minutiae.
5) Ehnancement To better highlight the edges of the crests, a Gabor filter is applied, whose parameters are set to take into account the information contained in the two vector and scalar fields, on the ϕ direction and on the previously

obtained crest frequency. After the enhancement it can be seen that the image is sharper, the contrast increases and the crests are better defined. However, areas that are too noisy often become a single dark spot and the broken ridges are not always recomposed.

6) Binarization This operator very simply transforms the gray-scale image obtained from enhancement into a black and white one. Discrimination occurs according to whether the pixel intensity is above or below a preset threshold. Normally this threshold is exactly in the middle of the scale (with gray levels from 0 to 255, the threshold is 128). The binarization, which in effect eliminates some information from the image, is however necessary to simplify the next operation.
7) Thinning The purpose of this operator is to thin all the crests present in the binarized image until they are reduced to the thickness of a single pixel. In this way, while preserving the general structure, a skeleton of the footprint is obtained which is simplified and more easily manageable by the algorithm that extracts the minutiae.
8) Minutiae extraction and filtering. This operator cycles through all the pixels one at a time, excluding the outer edges, then verifies that the point in question has not been previously excluded from the mask. If it passes this check, check that it is also full; if it is, it checks the eight adjacent pixels and counts how many of them are full. A typical fingerprint result is reported in Fig. 4 where only bifurcation and termination minutiae are considered. The two fingerprint images shown in this figure are related to the same finger and same individual (em).

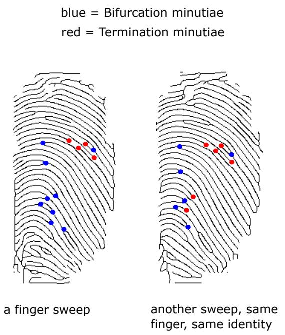
blue = Bifurcation minutiae
red = Termination minutiae
a finger sweep
another sweep, same finger, same identity

Fig. 4. Typical example of two fingerprints acquired with the sweep-type sensor and processed by the points described in this section.

A. Fingerprint Matching

The sweep-type fingerprint scanner we used allows us to simplify the fingerprint matching. Such operator will be described in this Section. The fingerprint which will be used for comparison as template is stored and the fingerprint to be verified is compared with the template. Representing the fingerprints with their minutiae, the patterns to be compared are the sequences of minutiae S (from "Stored") and V (from "to be Verified").

$$\begin{aligned} S &= \{M_1^S, M_2^S, \dots M_{N^S}^S\} \\ V &= \{M_1^V, M_2^V, \dots M_{N^V}^V\} \end{aligned} \tag{9}$$

The minutiae M_i^S and M_i^S of the "stored" and "to be verified" fingerprints are represented by their coordinates: $M_i^S = (x_i^S, y_i^S)$, $M_i^V = (x_i^V, y_i^V)$ where, x_i, y_i are the X-Y coordinate of the i-th S and V minutia, respectively. Likewise N^S and N^V are the number of the S and V minutiae respectively. The simplification lies in the fact that the direction of the minutiae is the same for the S and V fingerprints and thus they do not need not be estimated. We compute the distance between S and V fingerprints, namely the score, using a Dynamic Programming algorithm: the two sequences of minutiae are represented by sequences of couples, namely (x_1^S, y_1^S), (x_2^S, y_2^S), …, $(x_{N^S}^S, y_{N^S}^S)$ for the template S and (x_1^V, y_1^V), (x_2^V, y_2^V), …, $(x_{N^V}^V, y_{N^V}^V)$ for the new fingerprint V. Comparing three sequences with dynamic programming means that a warping path between the two sequences is computed. The warping path is made by K points where the generic point k is a correspondence between two generic couples (x_i^V, y_i^V) and (x_j^V, y_j^V). From the two couples a distance is defined using the Euclidean distance

$$d(k) = [(x_{i(k)} - x_{j(k)})^2 + (y_{i(k)} - y_{j(k)})^2]^{1/2} \tag{10}$$

Executing the Dynamic Programming recursion the score between the two fingerprints is computed as:

$$score(S, V) = min\frac{\sum_{k=1}^{K} d(k)w(k)}{\sum_{k=1}^{K} w(k)} \tag{11}$$

where the minimum is evaluated over all the warping paths, $w(k)$ is a weighting factor and the denominator compensates for the length of the warping path.

Using the score of authorized (A), and the average score of non-authorized (NA) we have:

$$\Lambda_F = \frac{score(S, V)^A}{score(S, V)^{NA}} \tag{12}$$

where x is a segment of Speech and the S apex stands clearly for Speech. Equation (12) in logarithmic terms, becomes:

$$\overline{\Lambda_F} = \log \Lambda_S = \log score(S, V)^A - \log score(S, V)^{NA} \tag{13}$$

This term is the score used in the data fusion algorithms. This score will be very high if the x speech segment comes from the owner of the model λ_A^S, while it will be very low if x speech segment is pronounced by an impostor.

V. Data Fusion Algorithms

A. Weighted-sum Fusion

The normalized quantities $\overline{\Lambda_F}$ and $\overline{\Lambda_S}$ can be combined in many ways. One possibility is to integrate them using a weighted average, which is shown in equation 14

$$\Lambda_1 = \alpha\overline{\Lambda_F} + (1-\alpha)\overline{\Lambda_S} \tag{14}$$

According to this equation, an index is obtained as a result of the combination. This value should be then subjected to a threshold ϕ to accept or reject the declared identity. In other words: If $\Lambda_1 \geq \phi$, then the identity could be verified. Otherwise, if $\Lambda_1 < \phi$, the identity could be rejected. In the following we describe a simple way to calculate α and ϕ. In a training phase, we enter a sequence of observations by the authorized person, $O_1^F, O_2^F, \ldots, O_N^F$, and $O_1^S, O_2^S, \ldots, O_N^S$ as input to the system.

We will therefore have a sequence of $\overline{\Lambda_1^F}, \overline{\Lambda_2^F}, \ldots, \overline{\Lambda_N^F}$ values and a sequence of $\overline{\Lambda_1^S}, \overline{\Lambda_2^S}, \ldots, \overline{\Lambda_N^S}$ values. We then calculate the mean and variance of these sequences: $\mu(\overline{\Lambda^F})$, $\mu(\overline{\Lambda^S})$ and $\sigma^2(\overline{\Lambda^F})$, $\sigma^2(\overline{\Lambda^S})$ First of all, we establish that a good value for the threshold is the mean of Λ_1:

$$\phi = \alpha\mu(\overline{\Lambda^F}) + (1-\alpha)\mu(\overline{\Lambda^S}) \tag{15}$$

Furthermore, if an authorized person provides his / her fingerprint and vocal observations, the value of Λ_1 is high but exhibit some variability as measured by its variance. However, it is important to minimize the variance of Λ_1 to minimize the false positives and false negatives in the verification process.

The minimization of $\sigma^2(x)$ is easily obtained as follows: since

$$\sigma^2(\Lambda_1) = \alpha^2\sigma^2\Lambda(F) + (1-\alpha)^2\sigma^2(\Lambda^S) + 2\alpha(1-\alpha)Cov(\Lambda^F, \Lambda^S) \tag{16}$$

and that fingerprints and voice prints are statistically independent, we have

$$\sigma^2(x) = \alpha^2\sigma^2\Lambda(F) + (1-\alpha)^2\sigma^2(\Lambda^S) \tag{17}$$

The value of α is obtained by setting to zero the derivative of $\sigma^2(x)$ with respect to α. Therefore:

$$\frac{d\sigma^2(x)}{d\alpha} = 2\alpha\sigma^2(\Lambda^F) - 2(1-\alpha)\sigma^2(\Lambda^S) = 0 \tag{18}$$

Hence:

$$\alpha\sigma^2(\Lambda^F) = (1-\alpha)\sigma^2(\Lambda^S) \tag{19}$$

or:

$$\alpha = \frac{\sigma^2(\Lambda^S}{\sigma^2(\Lambda^F)} \tag{20}$$

It can easily be shown that $\sigma^2(x)$ actually reaches a minimum for the value of calculated in (20).

Therefore, the algorithm developed for biometry is the following:

1) Calculation of Λ^F and Λ^S
2) Calculation of $\mu(\Lambda^F)$, $\sigma^2(\Lambda^F)$ and $\mu(\Lambda^S)$, $\sigma^2(\Lambda^S)$.
3) Calculation of $\Lambda_1 = \alpha\Lambda^F + (1-\alpha)\Lambda^S$ and $\phi = \alpha\mu(\Lambda^{(}F) + (1-\alpha)\mu(\Lambda^S)$

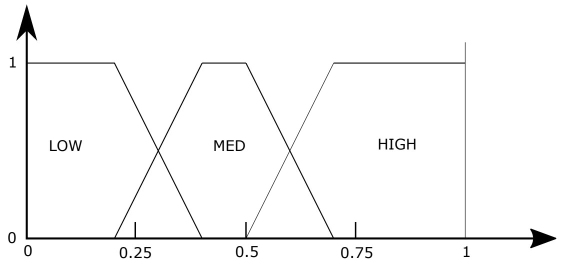

Fig. 5. Fuzzy sets

R1:	IF x1 is High AND x2 is High THEN u is High
R2:	IF x1 is Med AND x2 is Med THEN u is Med
R3:	IF x1 is Low AND x2 is Low THEN u is Low
R4:	IF x1 is High AND x2 is Low THEN u is Med
R5:	IF x1 is Low AND 2 is High THEN u is Med
R6:	IF x1 is Med AND x2 is High THEN u is Med
R7:	IF x1 is Med AND x2 is Low THEN u is Low
R8:	IF x1 is High AND x2 is Med THEN u is Med
R9:	IF x1 is Low AND x2 is Med THEN u is Low

TABLE I
FUZZY RULES

B. Fuzzy Fusion

Fuzzy fusion starts from the Λ^F and Λ^S measures described above.

These two indices are linearly normalized in the interval [0 - 1] by evaluating the minimum and maximum as indicated below. In case of the biometric system based on the voiceprint, the result has undergone a normalization on the interval [-0.8, 1.5], while, in the case of the biometric system based on fingerprint, normalization has been done on the interval [-2.24, 1.56]. Hence, it is possible to associate both results with the same fuzzy system. The normalized indices are provided to the fuzzy system defined in the next few paragraphs. Fig. 5 shows the fuzzy sets we used.

The Center method was used for the defuzzification method.

1) Fuzzy Rules: First, the system inputs are assigned: $x1 = \Lambda^S$ and $x2 = \Lambda^F$. These inputs are normalized between 0 and 1 and applied to the set of rules to find the fuzzy sets of the output variable $u = \Lambda_2$. The numerical value of the output is computed by applying the centroid method. The fuzzy rules are reported in Table I.

The output ranges from 0 and 1 so, in this case, the threshold can be set simply to 0.5.

2) Non-linear combination function: The fuzzy system created and described in the previous paragraphs builds one non-linear function with which the two inputs are integrated: $u = F(x1, x2)$. The form of this function is shown in the following figure.

VI. Experimental Results

Since no publicy datasets for voice and fingerprint biometric pairs is available, we developed our own dataset as described next in this section. Let us now first describe the speaker verification operations. Some operations were performed with the help of the tools provided by the ALIZELIA_RAL Speaker Verification Toolkit [20], [21].

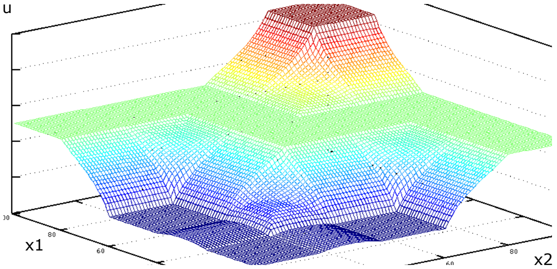

Fig. 6. Obtained non-linear integration function.

For the training and testing phases of the identity verification algorithm we have developed a dataset made up of vocal samples from 50 students, 15 females and 35 males, average age of 28 years. In particular, the subjects were required to read a series of 30 words 4 times each word. The acquired data is used for the first training phase and for the subsequent testing phase. The audio files were recorded with a slight background noise in order to evaluate the performance of the programs in the presence of a system that has noises (the average signal / noise ratio of the files is 15dB). The group of people is divided into 20 authorized and 30 unauthorized. For each of the 20 authorized, a model of each word is constructed using two repetitions of each word; the other two repetitions are used to obtain false positives in the test phase. Of the 30 not used, 20 are used to build the unauthorized model and 10 for the false negative test. Ultimately, the dataset we have created is made up of 6000 files. Of these, 1200 files are used for training and 1200 for false positive testing. In addition, 2400 files are used to build the unauthorized model and 1200 for the false negative test.

All 6000 files are converted into Cepstral Mel parameters using the 'bin/sfbcep' tool. Voice detection is provided by the 'bin/EnergyDetector' tool. The normalization of the Cepstral parameters is realized with the tool 'bin/NormFeat' while the model of the unauthorized with the tool 'bin/TrainWorld'. The 'bin/TrainTarget' tool is used for creating templates of authorized users. Finally, the tests are performed with the 'bin/ComputeTest' tool. The results related to identity verification through speaker verification are summarized in the following.

- Correct Positive Verification (CVP): They represent the cases in which a positive result of the verification is expected and obtained
- False Negatives (FRR): These represent the cases in which a positive result is expected and a negative result is obtained from the verification
- Correct Negative Verification (CVN): These represent the cases in which a negative verification result is expected and obtained
- False Positives (FAR): These represent the cases in which a negative result is expected and a positive feedback is obtained from the verification
- Accuracy: is the arithmetic mean of the CVP and CVN values

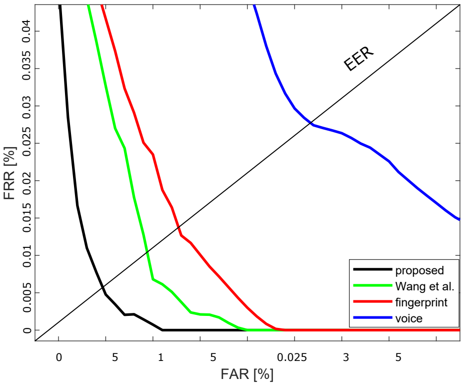

Fig. 7. Performance

Each authorized subject were also requested to sweep their selected finger 10 times. One fingerprint is used as a model of the the finger and the other 9 are used to test the false positives. Likewise each non authorized subject were requested to sweep the selected finger 20 times. Of the acquired fingerprints, 9 are used for measuring the false negatives, and 11 of them are used to form the non authorised subset. Fingerprint Identity verification has been performed using a series of tools developed by us in C++. The series of tools is summarized in Section 4. From speaker and fingerprint verification modules, the scores Λ_F and Λ_S are obtained. The next step is to integrate the scores obtained with voice and fingerprint the biometric systems The people who were used to conduct the tests using fingerprints are the same people used previously. it was therefore possible to carry out the tests of the overall system simply by associating each result obtained from the voice system with the corresponding result provided by the verification based on fingerprints. The data fusion algorithms, the conversion of $Lambda_1$ and $Lambda_2$ to likelihood and the Dempster-Shafer combination rule of the two belief functions are developed by us in C++ to obtain the final result.

The final results are reported in Fig.7, where ROC curves obtained with some identity verification algorithms are shown. The x axis represent the False Accept Rate (False Positives) while the y axis represents the False Reject Rate (False Negatives). Any point on the black line is an Equal Error Rate, namely a point where he false positive rate is equal to the false negative rate. The black curve is obtained with the algorithm proposed in this paper, which leads to an EER of about 0.005%. The green curve is obtained with the algorithm described by Wang *et al.* [16] where voice and fingerprint are fused with a User Specific Weighted Sum Rule, where the weight are selected specifically for each person. Using our dataset, its EER is about 0.1%. The red and blue curves are obtained with only fingerprint and only voice respectively. They lead to EER of about 0.015 and 0.025 respectively.

VII. Final Remarks and Conclusions

In this paper we present a bi-modal identity verification algorithm, based on voice and fingerprint biometrics. The voice biometric is processed using the classical GMM-UBM approach, and the fingerprint biometric is processed using the classical approach based on minutia extraction and matching. A matching matching approach based on dynamic programming is presented. The matching operation is simplified due to the sweep sensor used to acquire the fingerprint. One contribution of the paper is the Dempster-Shafer combination rule based fusion algorithm, which allows to combine different verification decisions coming from several data fusion algorithms. Verification results are performed using presented by ROC curves. Future works will evaluate the performance with greater dataset and the performance with other data fusion algorithms in addition to that used in this paper. Moreover we will study more thoroughly the conversion of the data fusion output values to likelihood values. Finally, we will investigate how our proposed framework can be made compliant with emerging challenges dictated by the novel big data trend (e.g., [22]–[31]).

Acknowledgements

This research has been partially supported by the French PIA project "Lorraine Université d'Excellence", reference ANR-15-IDEX-04-LUE.

References

[1] D. M. M. da Costa, H. dos Santos Passos, S. M. Peres, and C. A. M. Lima, "A comparative study of feature level fusion strategies for multimodal biometric systems based on face and iris," in *Proceedings of the annual conference on Brazilian Symposium on Information Systems, Information Systems: A Computer Socio-Technical Perspective, SBSI 2015, Goiânia, Brazil, May 26-29, 2015*. ACM, 2015, pp. 219–226.

[2] A. B. Khalifa and N. E. B. Amara, "Bimodal biometric verification with different fusion levels," in *Proceedings of the 6th International Multi-Conference on Systems, Signals and Devices*, 2009, pp. 1–6.

[3] Y. Li, G. Li, P. Li, S. Li, X. Yuan, D. Liu, and X. Yang, "A survey of multimodal fusion for identity verification," *Journal of Physics Conference Series*, pp. 1–5, 2020.

[4] S. A. and E. Rahman, "Multimodal biometric systems based on different fusion levels of ECG and fingerprint using different classifiers," *Soft Comput.*, vol. 24, no. 16, pp. 12 599–12 632, 2020.

[5] F. Ilkbahar and R. Kara, "Performance analysis of face recognition algorithms," in *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*, 2017, pp. 1–5.

[6] S. Madhavan and N. Kumar, "Incremental methods in face recognition: a survey," *Artif. Intell. Rev.*, vol. 54, no. 1, pp. 253–303, 2021.

[7] H. D. S. Gowda, G. H. Kumar, and M. Imran, "Robust multimodal biometric verification system based on face and fingerprint," in *2017 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2017, Udupi (Near Mangalore), India, September 13-16, 2017*. IEEE, 2017, pp. 243–247.

[8] N. Hassan, D. A. Ramli, and S. A. Suandi, "Fusion of face and fingerprint for robust personal verification system," *International Journal of Machine Learning and Computing*, vol. 4, pp. 371–375, 2014.

[9] M. F. R. Chowdhury, S. Selouani, and D. D. O'Shaughnessy, "Text-independent distributed speaker identification and verification using GMM-UBM speaker models for mobile communications," in *10th International Conference on Information Sciences, Signal Processing and their Applications, ISSPA 2010, Kuala Lumpur, Malaysia, 10-13 May, 2010*. IEEE, 2010, pp. 57–60.

[10] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *IEEE Trans. Speech Audio Process.*, vol. 19, no. 4, pp. 788–798, 2011.

[11] M. Rouvier, R. Dufour, and P. Bousquet, "Review of different robust x-vector extractors for speaker verification," in *28th European Signal Processing Conference, EUSIPCO 2020, Amsterdam, Netherlands, January 18-21, 2021*. IEEE, 2020, pp. 1–5.

[12] Y. Qian, Z. Chen, and S. Wang, "Audio-visual deep neural network for robust person verification," *IEEE ACM Trans. Audio Speech Lang. Process.*, vol. 29, pp. 1079–1092, 2021.

[13] P. Byahatti and M. S. Shettar, "Fusion strategies for multimodal biometric system using face and voice cues," *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 925, pp. 1–9, 2020.

[14] X. Zhang, D. Cheng, P. Jia, Y. Dai, and X. Xu, "An efficient android-based multimodal biometric authentication system with face and voice," *IEEE Access*, vol. 8, pp. 102 757–102 772, 2020.

[15] Q. A. Memon, "Multi-layered multimodal biometric authentication for smartphone devices," *Int. J. Interact. Mob. Technol.*, vol. 14, no. 15, pp. 222–230, 2020.

[16] Y. Wang, Y. Wang, and T. Tan, "Combining fingerprint and voiceprint biometrics for identity verification: an experimental comparison," in *Biometric Authentication, First International Conference, ICBA 2004, Hong Kong, China, July 15-17, 2004, Proceedings*, ser. Lecture Notes in Computer Science, D. Zhang and A. K. Jain, Eds., vol. 3072. Springer, 2004, pp. 663–670.

[17] G. Shafer, "A mathematical theory of evidence turns 40," *Int. J. Approx. Reason.*, vol. 79, pp. 7–25, 2016.

[18] F. Bimbot, J. Bonastre, C. Fredouille, G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-Garcia, D. Petrovska-Delacrétaz, and D. A. Reynolds, "A tutorial on text-independent speaker verification," *EURASIP J. Adv. Signal Process.*, vol. 2004, no. 4, pp. 430–451, 2004.

[19] N. Yager and A. Amin, "Fingerprint verification based on minutiae features: a review," *Pattern Anal. Appl.*, vol. 7, no. 1, pp. 94–113, 2004.

[20] F. Bellomo, F. Beritelli, and E. Sciacca, "Robustness of forensic speaker verification systems based on alizelia_ral toolkit," in *IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BIOMS)*. IEEE, 2014, pp. 1–6.

[21] J. F. Bonastre, F. Wils, and S. Meignier, "Alize, a free toolkit for speaker recognition," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '05, Philadelphia, Pennsylvania, USA, March 18-23*. IEEE, 2005, pp. 737–740.

[22] A. Campan, A. Cuzzocrea, and T. M. Truta, "Fighting fake news spread in online social networks: Actual trends and future research directions," in *2017 IEEE International Conference on Big Data (Big Data)*, 2017, pp. 4453–4457.

[23] M. Ceci, A. Cuzzocrea, and D. Malerba, "Effectively and efficiently supporting roll-up and drill-down olap operations over continuous dimensions via hierarchical clustering," *Journal of Intelligent Information Systems*, vol. 44, no. 3, pp. 309–333, 2015.

[24] A. Cuzzocrea and I.-Y. Song, "Big graph analytics: the state of the art and future research agenda," in *Proceedings of the 17th International Workshop on Data Warehousing and OLAP*, 2014, pp. 99–101.

[25] A. Cuzzocrea, "Accuracy control in compressed multidimensional data cubes for quality of answer-based olap tools," in *18th International Conference on Scientific and Statistical Database Management (SSDBM'06)*, 2006, pp. 301–310.

[26] A. Cuzzocrea and W. Wang, "Approximate range–sum query answering on data cubes with probabilistic guarantees," *Journal of Intelligent Information Systems*, vol. 28, no. 2, pp. 161–197, 2007.

[27] A. Bonifati and A. Cuzzocrea, "Storing and retrieving xpath fragments in structured p2p networks," *Data & Knowledge Engineering*, vol. 59, no. 2, pp. 247–269, 2006.

[28] M. Cannataro, A. Cuzzocrea, C. Mastroianni, R. Ortale, and A. Pugliese, "Modeling adaptive hypermedia with an object-oriented approach and xml," in *WebDyn@ WWW*, 2002.

[29] J. Fiérrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez, and J. Bigün, "Discriminative multimodal biometric authentication based on quality measures," *Pattern Recognit.*, vol. 38, no. 5, pp. 777–779, 2005.

[30] M. Tistarelli and J. Bigün, "Image and vision computing journal special issue on multimodal biometrics," *Image Vis. Comput.*, vol. 27, no. 3, p. 221, 2009.

[31] X. Li and C. Liu, "Big data biometrics processing: A case study of an iris matching algorithm on intel xeon phi," in *Big Data - Algorithms, Analytics, and Applications*, K. Li, H. Jiang, L. T. Yang, and A. Cuzzocrea, Eds. Chapman and Hall/CRC, 2015, pp. 393–404.