Advanced Multimodal Fusion for Biometric Recognition System based on Performance Comparison of SVM and ANN Techniques

Mofdi Dhouib
ATMS research unit,
National school of engineering,
Sfax, 3038, Tunisia
Higher Institute of Technological Studies
Department of Information Technologies.
El Bousten BP 88A, Sfax, 3099, Tunisia

Sabour Masmoudi
ATMS research unit,
National school of engineering,
Sfax, 3038, Tunisia
Ahmed Ben Hamida
ATMS research unit,
National school of engineering,
Sfax, 3038, Tunisia

ABSTRACT
Multimodal fusion for biometrics recognition system had gained specific attention nowadays thanks to its remarkable valuable results. For this approach, classification methods have been the basis of important recognition accuracy improvements. The artificial neural networks (ANN) and support vector machines (SVM) belong to this class of methods. This paper presents comparison concerning the performances of the some methods that have been successfully applied to the fusion of scores for multimodal biometric recognition. After recognizing each single modality which was the fingerprint, the face as well as the voice, we recovered three similarity scores. These scores are then introduced into the classification system based on neural networks and on support vector machine techniques. Experimental results demonstrate that the identity established by such an integrated system is more reliable than the established identity by fingerprint recognition system, facial verification system and a voice verification system. Fusion phases are performed at score level. An average rate (= 57.69 %) is obtained by fusion with ANN. While fusion with the SVM gives an average rate equal to (= 63.31 %). A brief introduction is provided regarding the commonly used biometrics, including face, fingerprint and voice. Comparing Merger methods is made according to criteria of optimization of error rate.

General Terms
Security, Human Factors, Recognition, Verification, Identification, Image Processing

Keywords
Multimodal biometric system, Voice, Fingerprint, Face, Recognition, Score-level, Fusion, ANN, SVM.

1. INTRODUCTION
Nowadays, there is a strong demand for automatic and secure identity verification systems. Biometric identification is a new technology to solve this problem. The fingerprint, face and speech are among the most commonly used biometric features.

In the current environment, the computer security has become a research area of great importance. In particular devise a reliable, efficient and robust identification system is a priority task. The individual’s identification has become essential to ensure the safety of systems and organizations. Faced with this increasing load, several methods for Biometric recognition have been proposed like: speaker recognition, facial recognition, fingerprint, iris recognition, retina, the hand shape.

Several works on multimodal biometric systems has already been done in the literature. Dieckmann et al. [1] proposed a summary level fusion scheme: "2-of-3 approach" that integrates the movement of the lips, face, and voice based on the principle that man uses, parallel, several indices identify a person, and Brunelli Falavian [2] proposed a system level measurement to combine the outputs of the sub-graders, Kitter et al. [3] demonstrated the effectiveness of an integration strategy that merges multiple snapshots of a biometric property initials using a Bayesian framework. Bigun et al. [4] proposed a Bayesian integration scheme of combining different evidence. Maes et al. [5] proposed to combine biometric data (e.g. fingerprint) with non biometric data (e.g. passwords). Hong and Jain [6], have developed a multimodal identification system that incorporates two different biometrics (fingerprints and face) that complement each other.

The unimodal biometric systems have been around for a few years but they are rather suitable for a medium security level. In fact, the higher the security level, the higher they tend towards the use of multimodal systems, more efficient and safer. In addition, systems based on a single biometric modality are vulnerable to attack. For the moment, no biometric indicator is 100% reliable according to [7]. This gave birth to the fusion of multimodal biometrics city before all the arguments over the results of various studies [8] [9] showed the performance of Multimodal Biometric systems over single-mode systems is a strong reason that we led to work on this topic (Add a term to a biometric system is to add a new source of information [8]).

The goal of this work is to provide a multimodal biometric system respecting several constraints comfort [10] and reliabilities (Increase rate recognition calculation inexpensive, robustness). In this context fusion allows address the lack of information resulting from the use of a single modality. This paper proposes an adaptive system of recognition of individuals by the merger of three biometric modalities: fingerprints, face and voice. Fusion was made using a hand
machines support vector (SVM), and artificial neural networks (ANN) on the other hand. These classification methods have greatly enriched the biometric recognition methods. Finally the results are compared.

The remaining of this article is organized as follows: Section II describes the unimodal biometric systems. Section III presents the proposed multimodal system using respectively ANN and SVM. Section IV discusses the experimental results of these approaches. The performance of the proposed multimodal approach using ANN is analyzed and compared with respect to that of the proposed multimodal approach using SVM. The final section, section V presents the conclusion and discusses perspectives of this work.

2. UNIODAL RECOGNITION

2.1. Fingerprint Recognition

This method relies on the principle of extracting the minutiae; settings relevant characteristic footprint such as Bifurcation: the point where the ridge is divided into two (Figure 1-a) and Ridge ending: the point where the ridge is stopped (Figure 1-b).

![Figure 1: Fingerprint minutiae](image)

The preprocessing phase is essential in a system for recognizing forms. To improve the quality of the information extracted from the images, one can specify regions of interest or enhance the contrast of images [5]. To avoid the extraction of false minutiae, several pretreatment steps have been performed like: Binarization, Skeletonization, (Thinning), Region of Interest, Minutiae extraction. The overall architecture of a fingerprint recognition system is described on figure 2.

![Figure 2: Principle of a fingerprint recognition system](image)

2.2. Face recognition

Facial recognition is a task that humans naturally and effortlessly perform in their daily lives. It is one of the basic biometric technologies, took a share of more and more important in the field of research, this being due to rapid advances in technologies such as digital cameras, Internet and mobile devices, all associated with security needs constantly increasing.

Facial recognition has several advantages over other biometric technologies. It is an inexpensive used technique, very well accepted by the public and requires no action by the user (Non-intrusive and no contact).

The basic principle of operation of a facial recognition system is illustrated by (Figure 3). It can be summarized in four stages: detection [3] and standardization [4] of the face and the last two represent the recognition made by a subsequent extraction phase a comparison of the characteristics.

![Figure 3: Principle of a facial recognition system](image)

2.3. Voice recognition

This is a transformation of a speech signal into a sequence of symbols representative of the signal content. The most commonly used extracting algorithms are the Mel frequency cepstral coefficients (MFCC) that showed on the following figure 4.

![Figure 4: Principle of the extraction of MFCC coefficients](image)
Although biometric recognition techniques can be very efficient, we cannot currently guarantee an excellent recognition rate with unimodal biometric systems based on a unique biometric signature.

Thus the error rate associated with unimodal biometric systems are relatively high, which makes them unacceptable for deployment of safety critical applications. To overcome these drawbacks, a solution proposed is the use of multiple biometric modalities in one system.

3. THE PROPOSED MULTIMODAL ARCHITECTURE

This proposed approach is to merge the output score of three different unimodal recognition systems use two types of classifiers. Then a performance comparison of the ANN merger with the SVM merger has been made.

3.1 Fusion with ANN

To achieve performance close to those observed in humans, the classifier based on artificial neural networks (ANN) have been used, associated with the fusion of the three modalities already cited. Indeed, using ANN for three separated biometrics, three different scores are recovered. These are supplied to a neural network composed of three classifiers to find the final score. Figure 5 shows the entire structure of the proposed system.

This proposed approach consists to fuse the output score of three different unimodal RNA. The fingerprint is combined with the face of a hand and the voice with the face of the other. Then the obtained output scores are combined to represent the input of a third network.

3.2 Fusion with SVM

There are two approaches for combining the separately and individual matching score. The first approach is to formulate it as a classification access, while the other approach is to treat it as a combination access.

The idea of the classification approach is to construct a feature vector using the matching scores output by the separate matchers. After that, this feature vector is classified into one of two classes: “Accept” (genuine user) or “Reject” (impostor). In general, the classifier utilized for this aim is able of acquiring knowledge of the decision frontier without regard for how the feature vector is constructed.

3.2.1 Overview of Support Vector Machine (SVM)

In 1992, Boser, Guyon, and Vapnik introduced Support Vector Machine (SVM) which became rather popular since SVM are a set of related supervised learning methods used for classification and regression [23]. They appertain to a family of generalized linear classifiers.

Vapnik have developed the foundations of Support Vector Machines (SVM) [24] which have been gained popularity due to many promising features such as better empirical performance. The formulation utilizes the Structural Risk Minimization (SRM) principle, which has been shown to be upper, to traditional Empirical Risk Minimization (ERM) principle, utilized by conventional neural networks. SRM minimizes a superior bound on the expected risk, whereas ERM keep down the error on the training data.

In biometrics, Support Vector Machine has been utilized for different learning based operations such as face recognition and multimodal fusion.

SVM is therefore a classifier that executes classification by building hyper planes in a multidimensional space and separating the data points into different classes.

3.2.2 Linearly separable data

Let \( \{ x_i, y_i \} \) be a set of \( N \) data vectors with \( x_i \in \mathbb{R}^n, y_i \in \{-1, 1\} \), and \( i = 1, \ldots, N \). \( x_i \) is the \( i \)-th data vector that belongs to a binary class \( y_i \).

A binary classifier should find a function \( f \) that maps the points from their data space to their label space

\[
 f : \mathbb{R}^n \rightarrow \{-1, 1\} \\
 x_i \rightarrow y_i 
\]

For the benefit of simplicity, we suppose that the data space is \( \mathbb{R}^2 \) and that a hyperplane separates the data. There are in fact an infinite number of hyperplanes that could divide the data into two classes. In accordance with the SRM principle, SVM
utilizes an iterative training algorithm which maximizes the margin between two classes to construct just one optimal hyperplane.

Assuming that we have a hyper plane separating the positive data and negative data, \( x_i \) belongs to the hyperplane which satisfies the relationship:

\[
\mathbf{w} \cdot x_i + b = 0
\]

In this equation \( w \) is the normal to the hyperplane and it is also a vector, \( b \) is the parameter of the hyperplane.

For mathematical calculations we have,

\[
w \times x_i + b = +1, \; y_i = +1
\]

\[
w \times x_i + b = -1, \; y_i = -1
\]

These equations can be combined in the following inequality:

\[
y_i (\mathbf{w} \times x_i + b) \geq 1
\]

The following figure shows the linearly separable case we have treated above:

![Figure 7: Linear separation hyperplane for linearly separable data.](image)

The points satisfying equality (2) belong to a hyper plane \( H_1 \):

\[
w \times x + b = +1
\]

Similarly, the checking point equal (3), belong to the hyperplane \( H_2 \):

\[
w \times x + b = -1
\]

The distance \( \mathbf{d} (\mathbf{w}, \mathbf{b}; \mathbf{x}) \) of a point \( \mathbf{x} \) from the hyperplane \( \mathbf{w} \times \mathbf{x} + \mathbf{b} = 0 \) is,

\[
d(\mathbf{w}, \mathbf{b}; \mathbf{x}) = \frac{|\mathbf{w} \mathbf{x} + \mathbf{b}|}{||\mathbf{w}||}
\]

Optimal hyper plane was constructed which the distance to the nearest points (margin) is Max. Maximize margin amounts to minimizing \( 2 / ||\mathbf{w}|| \). For this, the problem is reformulated as a dot product.

The support vectors and they belong to one of the hyperplanes \( H_1 \) or \( H_2 \). These points are closest to the border decision and they form the separator plan.

**B3. No linearly separable data**

If no hyperplane can be found to separate the data, a nonlinear mapping function is then needed. To overcome the disadvantages of non-linearly separable case, the idea of SVM is to change the data space. The data will be mapped nonlinearly in a high-dimensional space and the optimal hyper plane is computed in the high-dimensional space. The nonlinear transformation of data can allow linear separation examples in a new space. So we will have a change in dimension. This new dimension is called "re-description of space." Indeed, intuitively, the more the size of the re-description space, the greater the probability to find a separating hyper plane between examples is high. This is illustrated by the following scheme:

![Figure 8 : Non linearly separable data.](image)

Where examples are not linearly separable, the constraints (2) and (3) are released by introducing slack variables \( x_i \neq 0 \); \( i = 1 \ldots, l \) which become:

\[
w \times x_i + b = +1 - x_i, \; y_i = +1
\]

\[
w \times x_i + b = -1 + x_i, \; y_i = -1
\]

Therefore there is a transformation of a nonlinear problem of separation in the space of representation to a linear separation problem in an area of re-description of largest dimension. This non-linear transformation is performed using a specific kernel function.

**4. EXPERIMENTAL RESULTS**

The experimental results presented in this paper are divided into two parts. First the results obtained for each unimodal recognition system (fingerprint, face, and voice) are summarized. Secondly, the results of the proposed biometric multimodal fusion system used with three ANN classifiers are presented.

**4.1. Experimental results for the proposed architecture with ANN**

Table 2 summarizes the performance of the ANN fusion of the used biometric modalities.
Table 2: Performance of the modalities fusion

<table>
<thead>
<tr>
<th>Fusion of modalities</th>
<th>Number of epochs</th>
<th>HN 1000</th>
<th>5000</th>
<th>10000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voice/Face</td>
<td>5</td>
<td>18.39 %</td>
<td>20.11 %</td>
<td>22.14 %</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>21.55 %</td>
<td>28.7 %</td>
<td>31.55 %</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>34.60 %</td>
<td>40.37 %</td>
<td>43.75 %</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>44.9 %</td>
<td>48.03 %</td>
<td>56.40 %</td>
</tr>
<tr>
<td>Fingerprint/Face</td>
<td>5</td>
<td>15.03 %</td>
<td>21.69 %</td>
<td>27.15 %</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>23.90 %</td>
<td>28.00 %</td>
<td>34.29 %</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>35.85 %</td>
<td>37.12 %</td>
<td>43.65 %</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>42.01 %</td>
<td>45.2 %</td>
<td>54.8 %</td>
</tr>
<tr>
<td>Fingerprint/Face/Voice</td>
<td>5</td>
<td>11.00 %</td>
<td>18.50 %</td>
<td>27.87 %</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>22.30 %</td>
<td>28.32 %</td>
<td>35.61 %</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>35.85 %</td>
<td>41.64 %</td>
<td>48.30 %</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>43.97 %</td>
<td>48.2 %</td>
<td>57.69 %</td>
</tr>
</tbody>
</table>

From the Table 2, we can notice that the recognition rate is improved by the third classifier. The recognition rate is not enough to evaluate the performance of a biometric system. So the following table summarizes values of other performance criteria.

Table 3: Performance evaluation of fusion system using three ANN

<table>
<thead>
<tr>
<th>Fusion by three ANN</th>
<th>FRR</th>
<th>FAR</th>
<th>EER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fusion by three ANN</td>
<td>1.54 %</td>
<td>4.589 %</td>
<td>4.149 %</td>
</tr>
</tbody>
</table>

With: FRR is False Rejection Rate; FAR is Acceptance Refuse Rate and EER is Equal Error Rate.

4.2. Experimental results for the proposed architecture with SVM

Table 4: Performance evaluation of fusion system using SVM

<table>
<thead>
<tr>
<th>Fusion by linear kernel</th>
<th>FRR</th>
<th>FAR</th>
<th>EER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fusion by linear kernel</td>
<td>1.63 %</td>
<td>4.71 %</td>
<td>2.8158 %</td>
</tr>
<tr>
<td>Fusion by polynomial kernel</td>
<td>1.48 %</td>
<td>4.52 %</td>
<td>2.3467 %</td>
</tr>
</tbody>
</table>

Figure 9: FAR and FRR based on threshold (linear method).

Figure 10: FAR and FRR based on threshold (polynomial method).

Figure 11 shows the ROC curves and EER of the following biometric system: only face verification, only fingerprint verification, only voice verification and the proposed multimodal verification.

Multimodal biometric authentication based on score level fusion using SVM 5
Biometric systems are introduced. The principle is to design unimodal recognition systems and combine their scores from different biometric modalities to increase the power of identification.

The errors come from the imperfection of one biometric have been remedied by the fusion process by ensuring better recognition rate.

In addition, the concept of classification by neural network and support vector machines for multimodal fusion are detailed. Among the various levels of existing fusion, the score level is chosen because it offers the best compromise between the wealth of information and the ease of implementation.

This work provides new contribution to the field of biometrics multimodal. In fact, it shows the authentication of individuals by multimodal fusion based on ANN and SVM using the fingerprint, face and speech recognition.

The experimental results showed a significant improvement of SVMs compared to ANNs. This is due to what they can suffer multiple local minima, the solution to an SVM is global and unique. Two other advantages of SVMs are that it has a simple geometric interpretation and give a sparse solution.

From experiment results we obtain the following conclusions:

- The accuracy of verification is more improved than simply unimodal biometrics using the fusion of three biometric modalities.
- By comparing the results of SVM using a linear kernel with those using a non-linear kernel, we note an advantage of non-linear kernels. This is due that convexity is an interesting and important property of nonlinear SVM classifiers.
- Unlike SVMs computational complexity ANNs is proportional to the dimensionality of the input space. ANNs empirical use of risk minimization, while SVMs using structural risk minimization. Why SVMs outperform ANNs often in practice is that they deal with the biggest problem with ANNs, SVMs are less prone to overfitting.
- This method has the superiority over the previous methods due to the application of the new recognition algorithms and the SVM-based fusion rule.

Future work will investigate on better alternative recognition technique suitable for fusion of fingerprint, speech and face. The performance of multi-biometric systems can be improved if a suitable fusion strategy is used in particular for the system running an uncontrolled environment. Therefore, it would be interesting to apply other approach of fusion and to compare its results with those obtained by the ANN and SVM to maximize the performance of multi-biometric system.

6. REFERENCES


[8] Poinisot, Yang, fusion de biométries sans contact paume et visage Article, Université de Bourgogne, 2008.


[14] Thibaud, Réseaux de neurones en cascade pour la localisation précise de points caractéristiques du visage Article Université Pierre et Marie Curie. 2008


[17] Chaudhary, multimodal recognition system based on
fusion of Palm print, fingerprint and face, International Conference IEEE 2009

[18] Chollet, vérification biométrique multimodal: Le projet incitatif GET- BIOMET Telecom paris.2010


