On the use of genuine-impostor statistical information for score fusion in multimodal biometrics

Pascual EJARQUE*, Ainara GARDE*, Jan ANGUITA*, Javier HERNANDO*

Abstract

Matching score level fusion techniques in multimodal person verification conventionally use global score statistics in the normalization and fusion stages. In this paper, novel normalization and fusion methods are presented to take advantage of the separate statistics of the monomodal scores in order to reduce the genuine and impostor PDF lobe overlapping and improve the verification rate. Joint mean normalization is an affine transformation that normalizes the mean of the monomodal biometrics scores separately for the genuine and impostor individuals. Histogram equalization is used to align the statistical distribution of the monomodal scores and make the whole separate statistics comparable. The presented weighting fusion methods have been designed to minimize the variances of the separate multimodal statistics and reduce the multimodal PDF lobe overlapping. The results obtained in speech and face scores fusion upon POLYCOSt and XM2VTS databases show that the proposed techniques provide better results than the conventional methods.

Key words: Biometrics, Comparative study, Mixed method, Statistical method, Histogram, Experimental study, Speaker recognition, Image recognition, Face, Data fusion.

SUR L'USAGE DE L'INFORMATION STATISTIQUE CLIENT-IMPOSTEUR POUR LA FUSION DES SCORES EN BIOMÉTRIE MULTIMODALE

Résumé

Les techniques de fusion au niveau des degrés de pertinence dans la vérification multimodale de personnes utilisent conventionnellement des statistiques globales de pertinence pour les étapes de normalisation et de fusion. Dans le présent article, de nouvelles méthodes de normalisation et de fusion sont présentées pour profiter des statistiques séparées des pertinences monomodales en vue de réduire la superposition des densités de probabilité de client et d'imposteur et d'améliorer le taux de vérification. La normalisation conjointe de la moyenne est une transformation affine qui normalise la moyenne des qualifications biométriques monomodales séparément pour les individus client et imposteur. L'égalisation de l'histogramme est utilisée pour aligner la distribution statistique des pertinences monomodales et peut rendre comparables les statistiques complètes séparées. Les présentes méthodes de fusion avec pondération on été conçues de façon à minimiser les variances des statis-

* TALP Research Center, Department of Signal Theory and Communications, Technical University of Catalonia Barcelona, Spain; {pascual, agarde, jan, javier} @gps.tsc.upc.edu

ANN. TÉLÉCOMMUN., 62, n° 1-2, 2007
tiques multimodales séparées et réduire la superposition des densités de probabilité multi-
modales. Les résultats obtenus dans la fusion de pertinences pour voix et visage avec les
bases de données POLYCOAST et XM2VTS démontrent que la normalisation proposée et les tech-
niques de fusion produisent de meilleurs résultats que les méthodes conventionnelles.

Mots clés: Biométrie, Étude comparative, Méthode mixte, Méthode statistique, Histogramme, Étude expé-
imentale, Reconnaissance locuteur, Reconnaissance image, Visage, Fusion d’informations.

Contents

I. Introduction
II. Normalization methods
III. Fusion methods

IV. Recognition experiments
V. Conclusions

References (22 ref.)

I. INTRODUCTION

Multimodal recognition systems combine two or more of the human characteristics or
biometrics like voice [1], face [2], fingerprint [3], iris [4], hand geometry [5], etc. to achieve
better and more robust recognition results than using monomodal (one single) biometric
recognition systems [6]. In a multimodal recognition system, information can be integrated at
various levels: feature extraction level, matching score level and decision level. Fusion at the
feature extraction level combines different biometric features in the recognition process.
Score fusion matches the individual scores of different recognition systems to obtain a mul-
timodal score. Decision level systems perform logical operations upon the monomodal sys-
tem decisions to reach a final resolution. In this paper, novel matching score level fusion
techniques will be presented and compared with the most used conventional ones [7, 8].

A matching score level fusion system can be viewed as a two-steps process: normaliza-
tion and fusion itself [7, 8, 9, 10] as in figure 1. The normalization process converts the dif-
f erent scores in a comparable range of values. Some normalization efforts have been done for
monomodal systems. For example, Z-Norm and T-Norm [11] have been developed for
the normalization of speech scores and histogram equalization [12] is used for image enhan-
cement.

From a multimodal point of view, without a normalization of the scores, a biometric with
a higher range could eliminate the contribution of another with a lower one. Two conventio-
nal affine normalization techniques are: min-max, that linearly maps the scores in a range of
values between 0 and 1, and z-score, that transforms the scores to a distribution with zero
mean and unitary variance [7, 8]. Furthermore, Indovina et al. presented in [7] several non
affine adaptative normalizations as two-quadrics (QQ) and quadric-line-quadratic (QLQ) in order
to decrease the effect of the overlap of the genuine and impostor distributions.

In the fusion process the scores of the monomodal biometrics can be combined to obtain
a single multimodal score or can be classified in genuine or impostor in a verification process
or in the different persons in an identification process.

The most popular classificatory methods are Multilayer Perceptron (MLP) [13] and Sup-
port Vector Machines (SVM) [14]. The training phase of these techniques is made taking into
account the whole training data and not only its statistical properties. For this reason, they are called learning machines.

In the other hand, the most direct combinatory fusion methods are product and sum. In some fusion methods each biometric is weighted by a different factor, as in matcher weighting, where each monomodal score is weighted by a factor proportional to the inverse of the recognition result of the biometric, or in user weighting, where different weighting factors are applied for every user [7]. Other combinatory fusion methods are min-score and max-score that choose the minimum and the maximum of the monomodal scores [7, 8].

In [7], a user weighting method that depends on separate genuine and impostor statistics is presented by Indovina et al. Global normalization and fusion techniques can also be designed taking into account these separate statistics. In this sense, it can be seen that the variances of genuine and impostor scores give us a measure of the overlapping of the scores distribution lobes [15] and, in consequence, can give us an idea of the accuracy of a biometric technology. The use of this information in the score matching can improve the system performance.

The aim of this work is to present normalization and fusion methods considering separate distributions for the genuine and the impostor scores [16]. Firstly, we present the joint mean normalization method, which transforms each biometric in order to obtain the same statistical mean for the genuine and the impostor scores for all of them. Then, after joint mean normalization, all the normalized monomodal biometrics have the same defined mean for the genuine scores ($\mu_G$) and for the impostor scores ($\mu_I$). We show that, in certain conditions, using simple sum fusion after joint mean normalization, the multimodal biometric variances are reduced.

Secondly, we present cumulative distribution function equalization of the scores of the monomodal biometrics, also called histogram equalization, as a non affine normalization. As it has been said before, histogram equalization is a well-known method widely applied in the treatment of images in order to adjust their intensity histogram shape [12]. Cumulative distribution function equalization was mainly developed for the speech recognition adaptation approaches and for the correction of the non linear effects typically introduced by speech systems [17, 18]. In this work, this method is used for the comparison of the global histograms of the monomodal biometrics, which is expected to produce the equalization of the separate statistics for the genuine and impostor scores. We hope for a reduction of the multimodal separate variances by means of this separate statistics equalization.
Finally, two fusion methods are presented that weight the joint mean normalized biometrics in order to minimize the sum of the standard deviations or the sum of the variances of the genuine and imposter fusion scores. By minimizing the variance sum, the variances are reduced in all cases, independently of the monomodal statistical characteristics.

In section II, conventional normalization techniques, min-max and z-score, and two adaptative normalization techniques presented in previous works [7], two-quadrics (QQ) and quadric-line-quadric (QLQ), will we reviewed. Furthermore, we are going to introduce joint mean normalization and histogram equalization as normalization techniques.

In section III, three of the most usual fusion techniques, simple sum, matcher weighting, and SVM, will be described and two variance reduction weighting fusion techniques will be presented: minimum standard deviation sum weighting and minimum variance sum weighting.

In section IV, experimental results are described for the combination of speech and face scores. The scores have been obtained upon POLYCOST and XM2VTS databases respectively. In our experiments, the proposed methods have outperformed the conventional ones.

II. NORMALIZATION METHODS

In a multimodal fusion system, the normalization process transforms the monomodal scores in order to prepare them for the fusion process. When a global normalization is performed, that is, all the scores are treated by a monotonous increasing transformation, the performance of the monomodal system remains invariable. However, in the case of user dependant normalization methods, as Z-Norm or T-Norm, widely used in speech recognition systems, the performance of the monomodal systems change. In this paper, all the presented and tested normalization techniques are global and do not affect the performance of the monomodal systems.

II.1. Conventional normalization methods

In this section, the most usually used normalization methods, min-max and z-score, and two adaptative methods based in quadric shapes are reviewed.

Min-max (MM) normalization linearly maps the scores of a biometric in a range of values between zero and one. Equation (1) demonstrate the calculation of MM normalization where $x_{MM}$ is the min-max normalized biometric, $A$ is the original monomodal biometric, and $\min(A)$ and $\max(A)$ are the end points of the score range.

$$x_{MM} = \frac{a - \min(A)}{\max(A) - \min(A)}$$

Z-score (ZS) normalization sets the global mean of the scores of a monomodal biometric to zero and its global standard deviation and variance to one. Equation (2) must be applied for a z-score normalization where $x_{ZS}$ is the z-score normalized biometric, $A$ is the original
monomodal biometric, \( mean(A) \) is the statistical mean of the scores of \( a \) and \( std(A) \) is their standard deviation.

\[
x_{25} = \frac{a - mean(A)}{std(A)}
\]

Indovina et al. presented in [7] several adaptative normalization methods in order to reduce the overlap of the genuine and impostor lobes and, in consequence, increase the recognition system performance. These normalizations depend on the centre \((c)\) and the width of the overlap region \((w)\). Figure 2 shows the histogram of the genuine and impostor scores and the overlap region for a biometric.

![Figure 2](image_url)

**Fig 2.** - Genuine and impostor overlap for the biometric scores.

*Superposition des lobes client et imposteur des pertinences biométriques.*

Two of these adaptative methods, two-quadrics and quadric-line-quadric have been tested in order to compare them with the non affine normalization presented in this work: histogram equalization. These normalizations are applied upon the min-max normalized scores.

Two-quadrics \((QQ)\) is composed of two quadric segments that change the concavity at the centre of the overlap region \((c)\). The relationship among the normalized biometric scores \(x_{QQ}\) and the min-max normalized biometric scores \(x_{MM}\) is that in equation (3).

\[
x_{QQ} = \begin{cases} 
\frac{1}{c} x_{MM}^2 & x_{MM} \leq c \\
\frac{1}{c + \sqrt{(1 - c) (x_{MM} - c)}} & \text{otherwise}
\end{cases}
\]

In the Quadric-Line-Quadric \((QLQ)\) normalization the overlapped zone is left unchanged while the other regions are mapped with two quadric function segments. The relationship among the normalized biometric scores \(x_{QLQ}\) and the min-max normalized biometric scores \(x_{MM}\) is that in equation (4).
\[
(4) \quad x_{Q|Q} = \begin{cases} 
\frac{1}{(c - \frac{w}{2})} x_{MM}, & x_{MM} \leq \left( c - \frac{w}{2} \right) \\
\sqrt{\left(1 - c - \frac{w}{2}\right)} \frac{x_{MM} - c - \frac{w}{2}}{(c + \frac{w}{2})}, & (c - \frac{w}{2}) < x_{MM} \leq (c + \frac{w}{2}) \\
\text{otherwise} & 
\end{cases}
\]

II.2. Joint Mean Normalization (JMN)

As it has been said before any monotonous increasing transformation of the scores obtained with a person authentication system does not affect the system performance. Therefore, we can add a real constant to the biometric scores or we can multiply them by a positive real constant without changing the verification results. We assume the identity of the claimed and real persons known in the training phase. Let \( a_G \) and \( a_I \) be respectively the raw genuine and impostor scores for a monomodal biometric, and \( x_G \) and \( x_I \) the normalized scores computed as:

\[
(5) \quad x_G = k_1 \cdot a_G + k_2 \\
(6) \quad x_I = k_1 \cdot a_I + k_2
\]

where \( k_1 \) is a real positive constant and \( k_2 \) is a real constant. The scores \( x_G \) and \( x_I \) will yield to the same verification results than \( a_G \) and \( a_I \).

In the normalization step, the statistics of the genuine and the impostor scores are not generally taken into account. In this work, we propose to use this knowledge by normalizing the means of the two sets of scores. In this way, the absolute value of the means of \( x_G \) and \( x_I \) is set to a predefined value \( \mu_x = \mu_{xG} = -\mu_{xI} \) by adjusting the values of \( k_1 \) and \( k_2 \). In consequence, the sum of the genuine and impostor means will be zero. If \( \mu_{xG} \) and \( \mu_{xI} \) are set to one and minus one respectively, the obtained values of \( k_1 \) and \( k_2 \) are:

\[
(7) \quad k_1 = \frac{2}{\mu_{aG} - \mu_{aI}} \\
(8) \quad k_2 = \frac{\mu_{aG} - \mu_{aI}}{\mu_{aG} - \mu_{aI}}
\]

where \( \mu_{aG} \) and \( \mu_{aI} \) are the means of the genuine and impostor scores \( a_G \) and \( a_I \) respectively. The scores \( x_G \) and \( x_I \) will be referred to as joint mean normalized scores.

If this normalization is applied to two biometric set of scores \( a \) and \( b \), two joint mean normalized scores \( x \) and \( y \) are obtained:

\[
(9) \quad x = k_{1a} \cdot a + k_{2a} \\
(10) \quad y = k_{1b} \cdot b + k_{2b}
\]
where \( k_{1a}, k_{2a}, k_{1b} \) and \( k_{2b} \) can be calculated by means of equations (7) and (8) for each monomodal biometric.

Once the scores of the two biometric technologies are normalized, they have to be combined in order to obtain one single score and take a decision. One of the most straightforward fusion methods is simple sum (ss), which consists in the addition of the monomodal scores. However, in order to maintain the genuine and impostor means unchanged with respect to the monomodal systems, the half-sum will be applied in this work, i.e.:

(11) \[ u = \frac{1}{2} (x + y) \]

where \( u \) is the multimodal score. In this case, the means of the genuine and impostor scores of both the monomodal and multimodal systems are the same, i.e.:

(12) \[ \mu_{xG} = \mu_{yG} = \mu_{xI} = \mu_{yI} = \mu \]

where \( \mu_{xG}, \mu_{yG}, \mu_{xI}, \mu_{yI} \) are the means of the monomodal genuine and impostor scores \( x_G, y_G, x_I, y_I \) respectively. If we replace equations (9) and (10) in equation (11) we obtain the JMN-SS (Joint Mean Normalization – Simple Sum) scores:

(13) \[ u_{jmn-ss} = \frac{1}{2} \left( \frac{2}{\mu_{aG} - \mu_{al}} a - \frac{\mu_{aG} + \mu_{al}}{\mu_{bG} - \mu_{bl}} b - \frac{\mu_{bG} + \mu_{bl}}{\mu_{aG} - \mu_{al}} b \right) \]

where \( \mu_{bG} \) and \( \mu_{bl} \) are the means of the genuine and impostor scores \( b_G \) and \( b_I \) respectively.

As it has been explained above, the addition of a real constant or the multiplication by a positive real constant applied to a biometric score does not affect the performance of the system. In order to simplify the calculation of the scores of the fusion system, this property is applied to the JMN-SS scores to obtain a multimodal score \( v_{jmn-ss} \) that can be expressed as:

(14) \[ v_{jmn-ss} = \frac{a}{\mu_{aG} - \mu_{al}} + \frac{b}{\mu_{bG} - \mu_{bl}} \]

This result aims to calculate the multimodal biometric scores from the genuine and impostor scores and the statistical means of the monomodal scores \( a \) and \( b \).

At this point, the effect of the combination of JMN and ss upon the statistics of the multimodal biometric is described. The distribution of the genuine and impostor scores are overlapped and this overlap produces the verification errors. It can be expected that, if the variances of the genuine and the impostor scores are reduced, there will be a minor overlapping and, in consequence, the system will be improved. For this reason it is interesting to study the cases where the variances of the multimodal scores are reduced with respect to those of the monomodal ones. This study has only sense when the genuine and impostor means have a fixed value as in the case of JMN.

The variance of the genuine multimodal scores after JMN and half-sum can be calculated as [19]:

(15) \[ \sigma_{ug}^2 = E[(u_G - \mu_{uG})^2] = E \left[ \frac{1}{2} (x_G - \mu_{xG} + y_G - \mu_{yG}) \right]^2 \]

(16) \[ \sigma_{uG}^2 = \frac{1}{4} E \left[ (x_G - \mu_{xG})^2 + (y_G - \mu_{yG})^2 + 2 (x_G - \mu_{xG}) (y_G - \mu_{yG}) \right] \]
Finally, if we consider that \( x_G \) and \( y_G \), and \( x_I \) and \( y_I \) are uncorrelated and we apply the same result for the impostor scores, the variances for the genuine and impostor scores are

\[
\sigma_{uG}^2 = \frac{1}{4} (\sigma_{xG}^2 + \sigma_{yG}^2)
\]

\[
\sigma_{ui}^2 = \frac{1}{4} (\sigma_{xI}^2 + \sigma_{yI}^2)
\]

where \( \sigma_{xG}, \sigma_{yG}, \sigma_{xI} \) and \( \sigma_{yI} \) are the standard deviations of the genuine and impostor scores of the joint mean normalized monomodal systems.

Since we are looking for the reduction of the multimodal variance, we are going to find the relationship among the monomodal variances that will produce a reduction of the multimodal variances with respect to the monomodal ones. Then, for the genuine variances, it must be simultaneously accomplished this two equations:

\[
\sigma_{uG}^2 = \frac{1}{4} (\sigma_{xG}^2 + \sigma_{yG}^2) < \sigma_{xG}^2
\]

\[
\sigma_{ui}^2 = \frac{1}{4} (\sigma_{xG}^2 + \sigma_{yG}^2) < \sigma_{yG}^2
\]

The genuine variance of \( y \) can be expressed in function of the genuine variance of \( x \) as follows:

\[
\sigma_{yG}^2 = k \cdot \sigma_{xG}^2
\]

and the previous inequalities can be solved by replacing (21) in (19) and (20), i.e.:

\[
\frac{1}{4} (\sigma_{xG}^2 + k\sigma_{xG}^2) < \sigma_{xG}^2
\]

\[
\frac{1}{4} (\sigma_{xG}^2 + k\sigma_{xG}^2) < \sigma_{yG}^2
\]

Finally, we reach to the boundaries \( \sigma_{xG}^2 < 3\sigma_{xG}^2 \) and \( \sigma_{xG}^2 < 3\sigma_{xG}^2 \). In consequence, the genuine variance of the multimodal score is reduced when the genuine variance of each monomodal score is lower than three times the genuine variance of the other one. If we apply the same result for the impostor variances, the separate variances are reduced when

\[
\frac{1}{3} \sigma_{yG}^2 < \sigma_{xI}^2 < 3\sigma_{yG}^2
\]

\[
\frac{1}{3} \sigma_{yI}^2 < \sigma_{xI}^2 < 3\sigma_{yI}^2
\]

Obviously, the reduction of the monomodal variances implies a reduction of the multimodal variances. However, given a minimum variance value, the minimum values of \( \sigma_{uG} \) and \( \sigma_{ui} \) are achieved respectively when \( \sigma_{xG} = \sigma_{yG} \) and \( \sigma_{xI} = \sigma_{yI} \). In this case, the standard deviations are reduced by a \( \sqrt{2} \) factor.

As a result, if JMN and the half-sum combination rule is applied to uncorrelated biometric scores with similar variances, the genuine and impostor scores variances will be reduced and the verification results will probably improve with respect to those of the monomodal biometric systems.
II.3. Histogram Equalization (HEQ)

In the previous section we have reached the conclusion that the multimodal variances are reduced when the variances of the monomodal scores are similar. Unfortunately, we cannot expect this to be true in general. In order to solve this problem an equalization of the histograms of the monomodal scores is proposed in this paper as a non affine normalization process. Thus, the genuine and impostor statistics and, in consequence, the variances will probably be equalized.

Histogram equalization or cumulative distribution function equalization is a general non parametric method to make the cumulative distribution function (CDF) of some given data match to a reference distribution.

Histogram equalization is a widely used non linear method designed for the enhancement of images. Histogram equalization employs a monotonic, non-linear mapping which re-assigns the intensity values of pixels in the input image in order to control the shape of the output image intensity histogram to achieve a uniform distribution of intensities or to highlight certain intensity levels.

CDF equalization method was mainly developed for the speech recognition adaptation approaches or for the correction of non linear effects typically introduced by speech systems such as: microphones, amplifiers, clipping and boosting circuits and automatic gain control circuits. The principle of this method is to find a non linear transformation to reduce the mismatch of the statistics of two signals. By means of the cumulative distribution functions a transformation that maps the distribution of a signal back to the distribution of a reference signal is defined.

In the equalization process, the objective is to deal with more reliable data by performing cumulative distribution function estimation by intervals equally spaced out. Each interval \( x \in [q_i, q_{i+1}] \) is represented by \((x_i, F(x_i))\), that corresponds to the average of scores \(x_i\) and the maximum cumulative distribution value \(F(x_i)\), both calculated into each interval of the reference signal.

\[
x_i = \frac{\sum_{i=1}^{k_i} x_{ij}}{k_i}
\]

\[
F(x_i) = \frac{K_i}{M}
\]

where \( x_{ij} = x \in [q_i, q_{i+1}] \), \( k_i \) is the number of data in the interval \([q_i, q_{i+1}]\), \( K_i \) is the number of data in the interval \([q_0, q_{i+1}]\), and \( M \) is the total number of data.

\( F(x_i) \) defines the boundaries of the intervals in the CDF that will be equalized. These boundaries \([q'_i, q'_{i+1}]\), limit the interval of values that fulfils the following expression \( F(q'_i) \leq F(y) < F(q'_{i+1}) \). All values of \( y \) that are into the interval \([q'_i, q'_{i+1}]\) will be transformed to their corresponding \( x_i \) value.

In this work, the statistical matching technique matches the cumulative distribution function obtained from the speaker verification scores and the cumulative distribution function obtained from the face verification scores, both evaluated over the training data. The designed equalization takes as reference the histogram of the biometric with a better accuracy, which is expected to have lower separate variances, in order to obtain a bigger variance reduction.
In figure 3, several histograms of face and voice scores are plotted after the application of the presented normalizations in order to compare the transformations produced by each of them.

**Fig 3.** Scores distribution of face and speech biometrics after the presented normalizations.

*Distribution des pertinences pour voix et visage biométriques après les normalisations présentées.*

III. FUSION METHODS

III.1. Conventional fusion methods

In this section, we will review some usually used fusion methods including combinatorial and classification techniques.

Simple Sum (SS) is the most straightforward fusion method. The scores of the normalized monomodal biometrics are directly summed as shown in equation (30).

\[
(28) \quad u = x + y
\]

where \( x \) and \( y \) are the monomodal scores and \( u \) represents the multimodal scores.

Matcher weighting fusion (MW) makes use of the Equal Error Rate (EER). The weighting factor of each biometric is proportional to the inverse of its EER. If we want the sum of the weighting factors to be equal to one, the factors are calculated as

\[
(29) \quad u = \sum_{m=1}^{M} w^m x^m
\]

\[
(30) \quad w^m = \frac{1}{e^m} \frac{1}{\sum_{m=1}^{M} 1/e^m}
\]

where \( w^m \) and \( e^m \) are respectively the weighting factor and the EER for the \( m \)th biometric \( x^m \), and \( M \) is the number of biometrics. In this fusion method, the weights for more accurate matchers are higher than those of less accurate matchers.

Other fusion systems are based in classificatory techniques that are designed to decide at which class each experiment belongs to. The most usual techniques are Multi-Layer Perceptron (MLP) [13] and Support Vector Machines (SVM) [14]. In this work, SVMs have been employed to compare its results with those obtained with the novel techniques.

SVMs are kernel-based learning techniques that define hyperplanes in order to classify the experimental information. They are called learning methods because the hyperplane design is based in the experimental data and not in their statistical or in other characteristics. By means of a kernel the hyperplane design can be translated into a higher dimension space where the classification can be more accurate. Polynomial and radial basis functions (rbf) are the most used kernels. Furthermore, a user defined constant, C, can be set to control the misclassification of the experiments.

III.2. Linear weighting fusion for variance reduction

In the joint mean normalization section, the ss fusion method is used to obtain the multimodal score from the normalized monomodal scores. In this section, new fusion methods
that minimize the variance of the genuine and impostor multimodal scores are proposed. This new proposal is a linear combination of the monomodal scores which weights are chosen to minimize the variance of the combined value, i.e.

\[ u = \alpha \cdot x + (1 - \alpha) \cdot y \]  

where \( \alpha \) is a real positive number. The weighting factors \( \alpha \) and \( 1-\alpha \) guarantee that, if \( x \) and \( y \) have the same mean for the genuine and impostor scores, as in the case of joint mean normalized scores, \( u \) will also have the same means than \( x \) and \( y \).

\[
\begin{align*}
\sigma^2_{\mu_G} &= \alpha^2 \sigma^2_{xG} + (1 - \alpha)^2 \sigma^2_{yG} \\
\sigma^2_u &= \alpha^2 \sigma^2_{xl} + (1 - \alpha)^2 \sigma^2_{yl} 
\end{align*}
\]

Supposing that \( x_G \) and \( y_G \), and \( x_l \) and \( y_l \) are uncorrelated, the variances of \( u \) can be calculated as:

Unfortunately, in most cases there is no value of \( \alpha \) that minimizes both equations simultaneously. Therefore, in order to find one single value of \( \alpha \), a new criterion has to be defined. In this paper we propose to minimize the sum of the standard deviations or the sum of the variances instead of the variances separately. In the following sections these two options are considered.

III.2.1. Minimum standard deviation sum weighting (MDSW)

The objective of this fusion method is to find the value of \( \alpha \) in equation (31) that minimizes the sum of the standard deviations \( \sigma_{\mu_G} + \sigma_u \). If we derive this sum with respect to \( \alpha \) and equal the result to zero, we arrive to the following expression to obtain the variable \( \alpha \):

\[
\frac{\alpha^2 \sigma^2_{xG} + (1 - \alpha)^2 \sigma^2_{yG}}{\alpha^2 \sigma^2_{xl} + (1 - \alpha)^2 \sigma^2_{yl}} = \sqrt{\frac{\alpha^2 \sigma^2_{xG} + (1 - \alpha)^2 \sigma^2_{yG}}{\alpha^2 \sigma^2_{xl} + (1 - \alpha)^2 \sigma^2_{yl}}}
\]

The value of \( \alpha \) that minimizes the standard deviation sum fulfils this equation and, since there is not a closed expression to find \( \alpha \), it has to be solved by the use of iterative methods.

III.2.2. Minimum variance sum weighting (MVSW)

The sum of the variances \( \sigma^2_{\mu_G} + \sigma^2_u \) is minimized when its derivative with respect to \( \alpha \) is equal to zero, i.e.:

\[
\alpha \sigma^2_{xG} - (1 - \alpha) \sigma^2_{yG} + \alpha \sigma^2_{xl} - (1 - \alpha) \sigma^2_{yl} = 0
\]
The values of \( \alpha \) and \( 1-\alpha \) that minimize the sum of the variances are, in consequence:

\[
\alpha = \frac{\sigma_{yG}^2 + \sigma_{yI}^2}{\sigma_{xG}^2 + \sigma_{xI}^2 + \sigma_{yG}^2 + \sigma_{yI}^2}
\]

(40)

\[
1 - \alpha = \frac{\sigma_{xG}^2 + \sigma_{xI}^2}{\sigma_{xG}^2 + \sigma_{xI}^2 + \sigma_{yG}^2 + \sigma_{yI}^2}
\]

(41)

By the application of this technique every monomodal joint mean normalized biometric is multiplied by a factor proportional to the sum of the variances of the other biometric. Therefore, the most accurate biometry can be expected to have the lower sum of variances and, in consequence, the major weighting factor. However, a great difference among the sum of the variances of the monomodal biometrics produces weighting factors near to one and zero and, in consequence, the use of one single biometric.

In this case, the sum of the variances of the combined biometric is equal to the half of the harmonic mean of the sum of the variances of each monomodal biometric

\[
\sigma_{uG}^2 + \sigma_{ul}^2 = \frac{1}{\frac{1}{\sigma_{xG}^2 + \sigma_{xI}^2} + \frac{1}{\sigma_{yG}^2 + \sigma_{yI}^2}}
\]

(42)

If \( \sigma_{yG}^2 + \sigma_{yI}^2 \) is zero then \( \sigma_{uG}^2 + \sigma_{ul}^2 = \sigma_{xG}^2 + \sigma_{xI}^2 \). Any other value of \( \sigma_{yG}^2 + \sigma_{yI}^2 \) will increase the denominator of the expression and will decrease the value of the sum of the multimodal variances, that will be, in all cases, lower or equal than the sum of the \( x \) genuine and imposter biometric variances. This argument is also valid for the \( y \) biometric variances. Therefore, since no variance will be zero, this relationship guarantees that the sum of the variances of the multimodal biometric will be lower than the sum of the variances of any of the monomodal systems.

\[
\sigma_{uG}^2 + \sigma_{ul}^2 < \sigma_{xG}^2 + \sigma_{xI}^2
\]

(43)

\[
\sigma_{uG}^2 + \sigma_{ul}^2 < \sigma_{yG}^2 + \sigma_{yI}^2
\]

(44)

Since there is a reduction of the sum of the variances, an improvement of the performance of the system can be expected.

If we simplify the weighting expression as in the case of joint mean normalization with half sum fusion, we can define a multimodal biometric \( v_{mvs} \) that provides the same performance than joint mean normalization with minimum variance sum fusion. The scores of this biometric can be expressed as

\[
v_{mvs} = \frac{\mu_{aG} - \mu_{aI}}{\sigma_{aG}^2 + \sigma_{aI}^2} a + \frac{\mu_{bG} - \mu_{bI}}{\sigma_{bG}^2 + \sigma_{bI}^2} b
\]

(45)

This expression only depends on the scores and the statistics of the monomodal biometrics.

The weighting factors in expression (45) have some resemblances with d-distance used by Indovina et al in [7] for user weighting. However, d-distance has been tested as a global weighting method in our fusion system without satisfactory results.
IV. RECOGNITION EXPERIMENTS

In this section, we present the speech and face recognition systems involved in the fusion experiments and the experimental results obtained with the methods proposed in this paper and some of the most conventional ones.

IV.1. Experimental setup

The results presented in this work have been obtained in the score fusion level of integration. For this purpose, the scores obtained from one speech recognition expert and a face recognition expert have been combined to create a chimerical database with 10823 users by the combination of 110 users of the POLYCUST database and 340 users of the XM2VTS database.

In order to obtain the speech scores we have used a text-dependant recognition system based in Hidden Markov Models. Speech signals are segmented into frames of 20 ms with a shift of 10 ms in order to extract mel-cepstrum parameters. We have used two different combinations of MFCC parameters: 20 mel-cepstrum parameters (MFCC20) and 60 parameters including the 20 previous ones, 20 first derivative parameters and 20 second derivate parameters (MFCC60). The recognition system consists in a 32 gaussians per model GMM.

The database for the speech recognition experiments is POLYCUST [20], a telephonic voice database with 8 kHz sample rate. This database contains 134 speakers. We have used an English phrase pronounced 10 times for every speaker.

The face recognition expert is based in the NMFFACES algorithm [21]. Facial recognition systems are based on the conceptualization that a face can be represented as a collection of sparsely distributed parts: eyes, nose, cheeks, mouth, etc. Non-negative matrix factorization is used in Tefas et al. work to yield sparse representation of localized features to represent the constituent facial parts over the face images. Non-negative matrix factorization (NMF) is an appearance-based face recognition technique based on the conventional component analysis techniques.

The database for the face recognition experiments is XM2VTS database [22] of the University of Surrey. That is a multimodal database consisting in face images, video sequences and speech recordings of 295 subjects. For these experiments we have only used the face images. There are four frontal face images for subject.

One set of recordings of both databases has been used to train the recognition experts. Normalization and fusion techniques were trained by means of the monomodal recognition scores obtained from a second recording set. Finally, the results presented in this work were obtained by the application of the designed fusion techniques upon the monomodal scores obtained from a third set of recordings.

From the face and the speech recognition experts, 1488 speech experiments (for each combination of parameters) and 33361 face experiments were available. The scores of all the users were divided in two groups, as it has been said, for training and testing. Due to the great number of needed experiments for a statistically adequate number of errors, it was necessary the combination of one user from one biometric with more than one user from the other biometric. However, we have tried to minimize the number of combinations.
By the combination of the monomodal scores, a total of 29480 multimodal experiments have been created to train the normalization and fusion techniques and 30040 have been made to test them. All the results that are presented in this section were obtained with the test scores.

Before the training of the normalization and fusion techniques, a 1% of the extreme genuine and impostor scores have been removed for both biometrics, in order to avoid the negative effect of outliers. Furthermore, the number of intervals has been set to 1000 in the histogram normalization of the scores and, for SVMs, a radio basis function has been used and the user dependant constant C has been set to 10. These parameters have been adjusted in order to obtain the best performance for these techniques.

Apart from recognition results, sum of the variances of the genuine and impostor scores are presented in this work in order to compare them with the performance of the fusion techniques. All the variances have been calculated upon the joint mean normalized monomodal or multimodal scores, with genuine scores mean equal to one and impostor scores mean equal to minus one, in order to make these variances comparable.

IV.2. Results

In table I, we present the Equal Error Rate (EER) results and the sum of the variances of the genuine and impostor scores for the monomodal speech and face recognition systems.

<p>| EER monomodale pour voix et visage et somme de variances de clients et imposteurs. |
|---------------------------------|---------------------------------|</p>
<table>
<thead>
<tr>
<th>EER</th>
<th>Sum of Variances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voice</td>
<td>MFCC20 5.096 %</td>
</tr>
<tr>
<td>Voice</td>
<td>MFCC60 2.670 %</td>
</tr>
<tr>
<td>Face</td>
<td>2.064 %</td>
</tr>
</tbody>
</table>

These results show the straight-forward relationship among the recognition rates and the variances of the monomodal scores. The voice system using 20 parameters obtains the worst recognition accuracy and has the bigger sum of variances. In contrast, the face recognition system obtains the best recognition accuracy and the lower sum of variances.

In the next tables we will present the EER obtained by the combination of the different normalization and fusion techniques applied upon the speech and face recognition scores.

In our work, we have compared three affine normalization methods: min-max (MM), z-score (ZS) and joint mean normalization (JMN), three non-affine normalization methods: two-quadrics (QQ), quadric-line-quadratic (QLQ) and histogram equalization (HEQ), and five fusion methods: simple sum (SS), matcher weighting (MW), minimum standard deviation sum weighting (MSDSW), minimum variance sum weighting (MVSW) and Support Vector Machines (SVM).
However, the results obtained with the QQ normalization have been in all cases worst than that obtained with QLQ normalization, and, for this reason, are not presented in this paper. Furthermore, not all the combinations of normalization and fusion techniques have been applied. The variance reduction methods, MSDSW and MVSW have only sense after joint mean normalization and SVM implicitly applies MM normalization and, for this reason, only the combinations with this technique are presented in this work.

Firstly, we compare the results obtained by all the techniques upon the 20 parameters speech recognition system and the face recognition system. In table II, the results obtained by the affine normalization methods and all the fusion methods are presented.

<table>
<thead>
<tr>
<th></th>
<th>MM</th>
<th>ZS</th>
<th>JMN</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS</td>
<td>0.776%</td>
<td>0.752%</td>
<td>0.849%</td>
</tr>
<tr>
<td>MW</td>
<td>1.099%</td>
<td>0.832%</td>
<td>0.789%</td>
</tr>
<tr>
<td>MSDSW</td>
<td>-</td>
<td>-</td>
<td>0.789%</td>
</tr>
<tr>
<td>MVSW</td>
<td>-</td>
<td>-</td>
<td>0.782%</td>
</tr>
<tr>
<td>SVM</td>
<td>0.672%</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

The best result is obtained by the SVM fusion. For the combinatory fusion methods differences lesser than a 5% in the number of errors are obtained by ZS-SS, MM-SS and by JMN in combination with the weighting techniques: MW, MSDSW and MVSW.

The non-affine normalization techniques have been applied upon the MM normalized scores because of the definition of QQ and QLQ and for simplicity in the case of histogram equalization. However, in order to apply the variance reduction fusion techniques, the monomodal scores have to be joint mean normalized. For this reason, in our tests, we have performed the combination of non-affine normalizations and affine normalization. Due to the properties of the non-affine transformations, the application of the min-max normalization has no effect upon them. Then, in Table III we present the results for QLQ, QLQ-ZS, QLQ-JMN, HEQ, HEQ-ZE, and HEQ-JMN for the combinatory fusion techniques.

<table>
<thead>
<tr>
<th></th>
<th>QLQ</th>
<th>QLQ-ZS</th>
<th>QLQ-JMN</th>
<th>HEQ</th>
<th>HEQ-ZS</th>
<th>HEQ-JMN</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS</td>
<td>0.796%</td>
<td>0.746%</td>
<td>0.819%</td>
<td>0.802%</td>
<td>0.759%</td>
<td>0.759%</td>
</tr>
<tr>
<td>MW</td>
<td>1.092%</td>
<td>0.789%</td>
<td>0.796%</td>
<td>0.716%</td>
<td>0.739%</td>
<td>0.739%</td>
</tr>
<tr>
<td>MSDSW</td>
<td>-</td>
<td>0.789%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.636%</td>
</tr>
<tr>
<td>MVSW</td>
<td>-</td>
<td>-</td>
<td>0.789%</td>
<td>-</td>
<td>-</td>
<td>0.636%</td>
</tr>
</tbody>
</table>

The best result is obtained by the combination of HEQ and JMN normalizations with the variance reduction methods, MSDSW and MVSW, which obtain a better result than SVM, a 15% improvement upon the conventional affine techniques and a 10% upon all the normalization
techniques with the tested combinatory fusion techniques. It can also be noticed that by the application of the QLQ normalization, EER reduction is achieved for all the combinations except for MM-SS; however, this reduction is lesser than a 1% for its best result, QLQ-ZS-SS. In the case of HEQ, an improvement is achieved for all the normalization techniques in combination with the weighting fusion methods and for JMN-SS, and has reached a 19% improvement for its best results, the HEQ-JMN normalization in combination with the MSDSW and MVSW fusion techniques. The improvement achieved with respect to the better monomodal biometric (face) is of a 69%.

At this point, we are going to compare some recognition results with the sum of the variances of their genuine and impostor scores. In particular, we will compare the results and the variances of the multimodal scores based in the JMN and HEQ-JMN normalization techniques.

**Table IV.** — EER and Sum of variances for JMN based methods (MFCC20 and face).

<table>
<thead>
<tr>
<th></th>
<th>JMN</th>
<th>JMN Sum of variances</th>
<th>HEQ-JMN</th>
<th>HEQ-JMN Sum of variances</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS</td>
<td>0.849 %</td>
<td>0.5149</td>
<td>0.759 %</td>
<td>0.4000</td>
</tr>
<tr>
<td>MW</td>
<td>0.789 %</td>
<td>0.4813</td>
<td>0.739 %</td>
<td>0.4496</td>
</tr>
<tr>
<td>MSDSW</td>
<td>0.789 %</td>
<td>0.4822</td>
<td>0.636 %</td>
<td>0.3992</td>
</tr>
<tr>
<td>MVSW</td>
<td>0.782 %</td>
<td>0.4793</td>
<td>0.636 %</td>
<td>0.3992</td>
</tr>
</tbody>
</table>

Upon the results in Table IV, a clear relationship among the sum of the variances of the genuine and impostor scores and the accuracy of a multimodal biometric can be noticed. Only in the HEQ-JMN-SS case, a lower sum of variances does not lead to a better accuracy. For the rest of the techniques, a lower sum of variances implies a lower EER. This relationship is presented in figure 4.

![EER vs Sum of genuine and impostor variances](image.png)

**Figure 4.** — Sum of variances and EER relation for JMN based methods (MFCC20 and face).

*Relation entre la somme des variances et EER pour méthodes basées en JMN (MFCC20 et visage).*
When 60 parameters for the voice system are used, a similar analysis of the results can be done, taking into account that, in this case, both monomodal verification systems have more similar accuracy.

The results obtained by the affine normalization techniques in combination with the fusion techniques are shown in Table V.

<table>
<thead>
<tr>
<th></th>
<th>MM</th>
<th>ZS</th>
<th>JMN</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS</td>
<td>0.383%</td>
<td>0.426%</td>
<td>0.429%</td>
</tr>
<tr>
<td>MW</td>
<td>0.393%</td>
<td>0.413%</td>
<td>0.419%</td>
</tr>
<tr>
<td>MSDSW</td>
<td>-</td>
<td>-</td>
<td>0.376%</td>
</tr>
<tr>
<td>MVSW</td>
<td>-</td>
<td>-</td>
<td>0.383%</td>
</tr>
<tr>
<td>SVM</td>
<td>0.359%</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

The best result is obtained by the SVM fusion technique. For the combinatorial systems JMN normalization combined with the MSDSW fusion technique is the system that obtains a greater performance. However, JMN-MVSW and the fusion techniques with MM normalization obtain similar values.

The results obtained by the combination of non-affine and the combinatorial fusion techniques are that in Table VI.

<table>
<thead>
<tr>
<th></th>
<th>QLQ</th>
<th>QLQ-ZS</th>
<th>QLQ-JMN</th>
<th>HEQ</th>
<th>HEQ-ZS</th>
<th>HEQ-JMN</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS</td>
<td>0.353%</td>
<td>0.390%</td>
<td>0.399%</td>
<td>0.369%</td>
<td>0.363%</td>
<td>0.363%</td>
</tr>
<tr>
<td>MW</td>
<td>0.363%</td>
<td>0.370%</td>
<td>0.390%</td>
<td>0.360%</td>
<td>0.346%</td>
<td>0.346%</td>
</tr>
<tr>
<td>MSDSW</td>
<td>-</td>
<td>-</td>
<td>0.353%</td>
<td>-</td>
<td>-</td>
<td>0.353%</td>
</tr>
<tr>
<td>MVSW</td>
<td>-</td>
<td>-</td>
<td>0.353%</td>
<td>-</td>
<td>-</td>
<td>0.353%</td>
</tr>
</tbody>
</table>

By the use of non-affine normalizations, the best result is obtained by the application of histogram equalization and z-score or joint mean normalizations, in combination with the matcher weighting fusion technique which outperforms that obtained by SVM. For this best result, the reduction of the equal error rate with respect to the affine normalization methods is by an 8%. All the results obtained by the application of non-affine normalizations have outperformed that obtained by the correspondent affine normalizations even though, in this case, there is a lower difference among the results obtained by the application of non-affine normalization techniques. The improvement achieved with respect to the better monomodal biometric (face) is of an 83%.

It can be noticed that the performances of the fusion systems are more similar when 60 parameters for the voice recognition system are used, that is, when both monomodal sys-
tems have a smaller difference among their recognition results. In this case, the normalization and weighting factors are more similar for the different normalization and fusion systems than that used when the monomodal biometrics have a more different accuracy.

Table VII resumes the results and the sum of the variances of the techniques that make use of the JMN and HEQ-JMN normalizations.

<table>
<thead>
<tr>
<th></th>
<th>JMN</th>
<th></th>
<th>JMN</th>
<th></th>
<th>HEQ-JMN</th>
<th></th>
<th>HEQ-JMN</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sum of variances</td>
<td></td>
<td></td>
<td>Sum of variances</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS</td>
<td>0.429 %</td>
<td>0.3655</td>
<td>ME</td>
<td>0.363 %</td>
<td>0.3676</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MW</td>
<td>0.419 %</td>
<td>0.3650</td>
<td>MS</td>
<td>0.346 %</td>
<td>0.3673</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSDSW</td>
<td>0.376 %</td>
<td>0.3662</td>
<td>TV</td>
<td>0.353 %</td>
<td>0.3673</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MVSW</td>
<td>0.383 %</td>
<td>0.3656</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In this case, it can not be established a direct relation among the sum of the variances and the obtained verification rates. However, the difference among the sum of the variances of the different methods is very small and the recognition results are also very close. It can be concluded that there exist a relation among the variances and the recognition result in a general manner, but it is not directly applicable when there are little differences among the variances.

Besides the results obtained by every fusion technique some other questions have to be taken into account in order to choose a fusion approach. To finish the results section a comparison of the advantages and disadvantages of each technique is made.

The application of the MM normalization is very simple due to it only depends on the extremes values of the scores. However, this technique can be affected by the presence of outliers. ZS and JMN depend on the means and variances of the scores and, for this reason, they are less affected by the outliers. The better results obtained by these three normalization techniques are similar.

The parameters needed for the application of QQ and QLQ normalization are the centre and the width of the overlap region of genuine and impostor training scores. Although some improvement is obtained by these techniques, outliers in the maximum impostor scores or in the minimum genuine scores can hardly affect these normalization techniques. HEQ takes into account an approximation to the whole statistical properties of the monomodal scores. In consequence, this technique is more resistant to the presence of outliers but the computational requirements are greater in the training phase. This normalization technique has obtained the best recognition results.

As for the fusion techniques, SS is the most straight-forward and simple technique and MW is based in the individual recognition result obtained by each monomodal biometric system. In other hand, the variance reduction methods, MSDSW and MVSW, take into account the reduction of the separate variances of the genuine and impostor scores. All four methods are easy to train and to apply in the test phase. The learning machines, as SVM, require more computational and memory resources, as much in the training phase as in the one of test. In this work, it has been demonstrated that the results obtained by a learning machine can be
outperformed by the use of the knowledge of the statistics and the performance of the mono-modal biometrics.

V. CONCLUSIONS

In the score matching level fusion landscape, we have presented normalization and fusion methods based in the separate treatment of genuine and impostor score statistics.

It has been demonstrated that, in some conditions, variance reduction can be achieved by the use of the presented joint mean normalization method in combination with simple sum fusion.

Histogram normalization or cumulative distribution function equalization has been presented as a normalization method. The designed equalization takes as a reference the histogram of the biometric with a better accuracy that can be expected to have lower separate variances, in order to obtain a bigger variance reduction.

Furthermore, variance minimization fusion methods have been presented: minimum standard deviation sum weighting and minimum variance sum weighting. This last method yields to a multimodal biometric with a lower sum of the genuine and impostor variances than that of the original monomodal biometrics.

JMN and the linear weighting fusion methods for variance reduction: MSDSW and MVSW improve or obtain similar results to that obtained by the conventional normalization and combinatorial fusion methods. Furthermore, the use of the histogram equalization as a normalization technique improves the results obtained by the conventional and other non-affine normalization techniques. In particular, the combination of HEQ, JMN and the weighting fusion techniques obtains the better results in this work including that obtained by a SVM.

In other hand, it has been probed a certain relationship among the variances of the genuine and impostor scores and the accuracy of a biometric. This relationship can be exploited in order to increase the multimodal biometrics accuracy.

Manuscrit reçu le 9 novembre 2005
Accepté le 28 juin 2006

Acknowledges

We want to thank Dr. A. Tefas who has provided us with face recognition results in order to test our normalization and fusion techniques.
REFERENCES


[22] Lütthin (J.), Maitre (G.), Evaluation Protocol for the Extended M2VTS Database (XM2VTSDB), IDIAP Communication 98-05 (1998), Martigny, Switzerland.