Variance reduction by using separate genuine-impostor statistics in multimodal biometrics

P. Ejarque, J. Hernando
TALP Research Center
Department of Signal Theory and Communications
Technical University of Catalonia, Barcelona, Spain
javier@gps.tsc.upc.edu

Abstract

In this paper, we propose some novel normalization and fusion techniques for biometric matching score level fusion in person verification. While conventional matching score level fusion methods use global score statistics, we consider in this work both genuine and impostor statistics separately. Performing a joint mean normalization of the separate monomodal scores, multimodal scores with less separate variance than the monomodal ones are obtained. Furthermore, a weighting method has been designed in order to minimize the variance sum of the separate multimodal statistics. This method obtains a minor sum of genuine and impostor variances for the multimodal biometric than that of the monomodal ones. The results obtained in speech and face scores fusion upon POLYCOST and XM2VTS databases show that the proposed normalization and fusion techniques provide better results than the conventional methods.

1. Introduction

Multimodal recognition systems combine two or more of the human characteristics or biometrics like voice, face, fingerprint, iris, hand geometry, etc. to achieve better and more robust recognition results than using monomodal (one single) biometric recognition systems [1]. 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 multimodal score. Decision level systems perform logical operations upon the monomodal system decisions to reach a final resolution. In this paper, novel matching score level systems will be presented and compared with the most used conventional ones.

A matching score level fusion system can be viewed as a two-steps process: normalization and fusion itself [2, 3, 4, 5]. The normalization process converts the different scores in a comparable range of values. Without this step, a biometric with a higher range could eliminate the contribution of another with a lower one. Two conventional 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 [2, 3].

On the other hand, most direct fusion methods are product and sum. Furthermore, some fusion methods weight each biometric using sum or product operations. This is the case for matcher weighting, that weights each monomodal score with the inverse of the recognition result of the biometric, and for user weighting, that applies different weighting factor for every user [2]. Other fusion methods are min-score and max-score that choose the minimum and the maximum of the monomodal scores as the multimodal score [2, 3].

In [2], 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 [6] and, in consequence, can give us an idea of the accuracy of a biometric. 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. Firstly, we present the joint mean normalization method, which transforms each biometric in order to obtain the same statistical means for the genuine and the impostor scores for all of them. We show that, in certain conditions, using simple sum fusion after joint mean normalization, the multimodal biometric variances can be reduced in relation to the monomodal ones.

Secondly, 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 genuine and impostor multimodal scores. By minimizing the variance sum the variances are reduced in all cases, independently of the monomodal statistical characteristics.

In section 4, 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 most used conventional ones.

2. Joint mean normalization and variance reduction

In a recognition system we can add a real constant to biometric scores or we can multiply them by a positive real constant without changing the sorting of the scores and, in consequence, without changing the system performance.

Let \( a_G \) and \( a_I \) be respectively the raw genuine and impostor scores for a monomodal biometric.

The scores \( x_G \) and \( x_I \) computed as

\[
    x_G = k_1 \cdot a_G + k_2 \quad x_I = k_1 \cdot a_I + k_2 \quad (1)
\]

where \( k_1 \) is a real positive constant and \( k_2 \) is a real constant will yield to the same recognition results than \( a_G \) and \( a_I \).
Then, we can adjust the values of $k_1$ and $k_2$ such as the absolute value of the mean of $x_G$ and $x_I$ are a defined value, that is

$$\mu_G = -\mu_I = \mu.$$  

The sum of the mean of the scores $x_G$ and $x_I$ is zero. In the following, they will be referred to as joint mean normalized scores.

If we apply this normalization to two biometrics $a$ and $b$ we obtain two joint mean normalized biometrics $x$ and $y$ respectively. Figure 1 plots the histogram of the joint mean normalized scores for a speech recognition system that uses mel-frequency cepstrum features (MFCC) over the POLYCOST database ($x$) and a face recognition system over the XM2VTS database ($y$).

![Figure 1: Joint mean normalized scores distribution of speech ($x$) and face ($y$) biometrics.](image)

Let us apply simple sum, that is the most straightforward fusion method, after this joint mean normalization. However, in order to maintain the mean unchanged among the monomodal biometrics and the resultant multimodal biometric, the half-sum case will be applied.

$$u = \frac{1}{2}(x + y)$$  

In this case,

$$\mu_G = -\mu_I = \mu_G = -\mu_G = -\mu_G = \mu.$$  

As we can see in figure 1, the lobes for genuine and impostor scores are overlapped. This overlap produces the recognition errors. It can be expected that if the genuine and the impostor scores variances are reduced, there will be a lesser overlapping and, in consequence, the system will be improved.

We will now calculate the standard deviations for the genuine and impostor multimodal scores and compare them with $x$ and $y$ standard deviations.

If we consider that $x_G$ and $y_G$ are uncorrelated as well as $x_I$ and $y_I$ it is well-know [7] that

$$\sigma_{xG}^2 = \frac{1}{4}(\sigma_{xG}^2 + \sigma_{yG}^2) \quad \sigma_{xI}^2 = \frac{1}{4}(\sigma_{xI}^2 + \sigma_{yI}^2)$$  

It is easy to demonstrate that the variances are reduced when

$$\frac{1}{3}\sigma_{xG}^2 < \sigma_{xI}^2 < 3\sigma_{xG}^2 \quad \frac{1}{3}\sigma_{xI}^2 < \sigma_{xG}^2 < 3\sigma_{xI}^2$$  

In consequence, a reduction in the variances is achieved when the variance of a biometric is 3 times lesser than the variance of the other one.

The minimum value of $\sigma_{xG}$ and $\sigma_{xI}$ are obtained when $\sigma_{xG} = \sigma_{yG}$ and $\sigma_{xI} = \sigma_{yI}$ respectively. Then, the standard deviations will be reduced by a factor of $\sqrt{2}$.

In conclusion, if we apply joint mean normalization and half-sum combination rules to uncorrelated biometrics with similar variances the genuine and impostor scores variances will be reduced and we could hope to improve the recognition results of both individual biometrics.

In order to reduce the complexity in the calculation of the scores of this fusion system and as we have seen above, we can add a real constant to biometric scores or we can multiply all them by a positive real constant without changing the recognition results. We can apply this property to the multimodal scores to obtain a new biometric $v_{\text{mne-b}}$ that can be expressed as

$$v_{\text{mne-b}} = \frac{a}{\mu_{xG} - \mu_{xI}} + \frac{b}{\mu_{yG} - \mu_{yI}}$$  

This result aims to calculate the multimodal biometric from the genuine and impostor means of $a$ and $b$.

### 3. Variance minimization using linear weighting fusion

In order to obtain a major standard deviation reduction than in the previous case we are going to calculate a multimodal biometric as the linear combination of the joint mean normalized biometrics, that is, we are going to weight each biometric in order to minimize standard deviations. The expression for the multimodal biometric scores is then

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

where $\alpha$ is a real positive numbers. The weighting factors $a$ and $(1-\alpha)$ guarantees the same mean for genuine and impostors scores of $u$, $x$ and $y$.

Then, supposing $x_G$ and $y_G$ are uncorrelated as well as $x_I$ and $y_I$ variances can be calculated as

$$\sigma_{xG}^2 = \alpha^2 \sigma_{xG}^2 + (1-\alpha)^2 \sigma_{yG}^2$$  

$$\sigma_{xI}^2 = \alpha^2 \sigma_{xI}^2 + (1-\alpha)^2 \sigma_{yI}^2$$  

In most cases, there is no value of $\alpha$ than could accomplish both equations simultaneously.

In order to find one single value of $\alpha$, we have to define new accuracy criteria for the combined biometric. Then, instead of minimizing the standard deviations separately we are going to minimize the sum of these standard deviations or the sum of the variances.

#### 3.1. Minimum standard deviation sum

In this section, we are going to find the value of $\alpha$ that minimize $\sigma_{xG}^2 + \sigma_{xI}^2$.

If we derive this sum and equal the result to zero, we arrive to the following expression to obtain the variable $\alpha$
that minimizes the standard deviation sum.

### 3.2. Minimum variance sum

The sum of the variances is minimized for the following value of $\alpha$

$$\alpha = \frac{\sigma^2_{\text{io}} + \sigma^2_{\text{ik}}}{\sigma^2_{\text{io}} + \sigma^2_{\text{ik}} + \sigma^2_{\text{io}} + \sigma^2_{\text{ik}}}$$  \hspace{1cm} (12)$$

In this case, the expression for the sum of the variances of the combined biometric is

$$\sigma^2_{\text{io}} + \sigma^2_{\text{ik}} = \frac{1}{\sigma^2_{\text{io}} + \sigma^2_{\text{ik}} + \frac{1}{\sigma^2_{\text{io}} + \sigma^2_{\text{ik}}}}$$  \hspace{1cm} (13)$$

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.

This relationship guarantees that the sum of the variances of the combined biometric will be lesser than the sum of the variances of all the monomodal biometrics.

$$\sigma^2_{\text{io}} + \sigma^2_{\text{ik}} < \sigma^2_{\text{io}} + \sigma^2_{\text{ik}} + \sigma^2_{\text{io}} + \sigma^2_{\text{ik}}$$ \hspace{1cm} (14)$$

We can expect then that this method will improve any of the monomodal biometrics. However, a great difference among variances sum of the monomodal biometrics produces, in fact, weighting factors near to 1 and 0, and, in consequence, the use of only one single biometric. In other cases, we can expect that the biometric with a major accuracy will have the lower sum of variances and, in consequence, the major weighting factor.

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

$$v_{\text{mv}} = \frac{\mu_{\text{io}} - \mu_{\text{ik}} + \mu_{\text{io}} - \mu_{\text{ik}}}{\sigma^2_{\text{io}} + \sigma^2_{\text{ik}}}$$  \hspace{1cm} (15)$$

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

The weighting factors in expression (15) have some resemblances with d-distance used by Indovina et al in [2] for user weighting. However, d-distance has been tested as a global weighting method in our fusion system without satisfactory results.

### 4. Recognition experiments

In this chapter, we will present the speech and face recognition systems used in the fusion experiments and the experimental results obtained with the methods proposed in this paper and the most used conventional ones.

#### 4.1. Experimental setup

To obtain the speech scores we have used a text-dependent recognition system based in Hidden Markov Models. We have used different combinations of MFCC parameters: 20 parameters (MFCC20), 40 parameters including the previous ones and 20 first derivative parameters (MFCC40) and 60 parameters including the 40 previous ones and 20 second derivative parameters (MFCC60).

The database used for the speech recognition experiments is POLYCOSt database [8], 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 database used for the face recognition experiments is XM2VTS database [9] of the University of Surrey. That is a multi-modal 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.

The fusion experiments combine the scores for the users of both recognition systems. A chimerical database with 10823 users has been created by combining 110 users of the POLYCOSt database and 340 users of the XM2VTS database. A total of 59520 experimental results have been used by the combination of 1488 speech experiments (for each combination of parameters) and 33631 face experiments.

#### 4.2. Results

In Table 1, we present the Equal Error Rate (EER) results of the individual speech and face recognition systems.

<table>
<thead>
<tr>
<th></th>
<th>EER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voice</td>
<td></td>
</tr>
<tr>
<td>MFCC20</td>
<td>5.702 %</td>
</tr>
<tr>
<td>MFCC40</td>
<td>3.614 %</td>
</tr>
<tr>
<td>MFCC60</td>
<td>2.670 %</td>
</tr>
<tr>
<td>Face</td>
<td>2.174 %</td>
</tr>
</tbody>
</table>

**Table 1**: Individual speech and face EER.

In the next tables we present the different EER obtained with the combination of the different normalizations and fusion techniques for every speech recognition system combined with the face recognition system.

We compare three normalization methods: min-max (MM), z-score (ZS) and joint mean normalization (JMN) and six fusion methods: simple sum (SS), matcher weighting (MW), that weights each score with the inverse of his recognition result, minimum (MIN) and maximum (MAX), that choose the minimum and maximum of the monomodal scores respectively, minimum standard deviation sum weighting (MSDSW) and minimum variance sum weighting (MVSW). These two last fusion methods have only sense with joint mean normalization.

<table>
<thead>
<tr>
<th></th>
<th>MM</th>
<th>ZS</th>
<th>JMN</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS</td>
<td>0.953 %</td>
<td>1.170 %</td>
<td>1.307 %</td>
</tr>
<tr>
<td>MW</td>
<td>1.262 %</td>
<td>0.966 %</td>
<td>0.939 %</td>
</tr>
<tr>
<td>MIN</td>
<td>2.740 %</td>
<td>2.740 %</td>
<td>2.740 %</td>
</tr>
<tr>
<td>MAX</td>
<td>1.823 %</td>
<td>1.415 %</td>
<td>1.546 %</td>
</tr>
<tr>
<td>MSDSW</td>
<td>-</td>
<td>-</td>
<td>0.966 %</td>
</tr>
<tr>
<td>MVSW</td>
<td>-</td>
<td>-</td>
<td>0.958 %</td>
</tr>
</tbody>
</table>

**Table 2**: MFCC20 and face fusion results.
4. Experimental Results

The experimental results obtained with the presented joint mean normalization (JMN) improve that obtained with the classical normalization techniques (SS and ZS). The best results have been obtained with the weighting fusion techniques: matcher weighting (MW), minimum standard deviations sum weighting (MSDSW) and minimum variance sum weighting (MVSW).

5. 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 shown 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.

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

The experimental results obtained with the presented joint mean normalization (JMN) improve that obtained with the classical normalization techniques (SS and ZS). The best results have been obtained with the weighting fusion techniques: matcher weighting (MW), minimum standard deviations sum weighting (MSDSW) and minimum variance sum weighting (MVSW).

6. Acknowledgements

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

The work has been partially supported by the EU-funded project CHIL (IP506909).

7. References