BIOMETRIC SOURCE WEIGHTING IN MULTI-BIOMETRIC FUSION: TOWARDS A GENERALIZED AND ROBUST SOLUTION

Naser Damer, Alexander Opel, Alexander Nouak

Fraunhofer Institute for Computer Graphics Research (IGD)
Darmstadt, Germany
Email: {naser.damer, alexander.opel, alexander.nouak}@igd.fraunhofer.de

ABSTRACT

This work presents a new weighting algorithm for biometric sources within a score-level multi-biometric system. Those weights are used in the effective and widely used weighted sum fusion rule to produce multi-biometric decisions. The presented solution is mainly based on the characteristic of the overlap region between the genuine and imposter scores distributions. It also integrates the performance of the biometric source represented by its equal error rate. This solution aims at avoiding the shortcomings of previously proposed solutions such as low generalization abilities and sensitiveness to outliers. The proposed solution is evaluated along with the state of the art and best practice techniques. The evaluation was performed on two databases, the Biometric Scores Set BSSR1 and the Extended Multi Modal Verification for Teleservices and Security applications database and a satisfying and stable performance was achieved.

***Index Terms*—** Multi-biometric fusion, Biometric source weighting, Score-level fusion.

1. INTRODUCTION

Multi-biometrics tries to use multiple biometric information sources to enhance performance and to overcome the limitations of the conventional uni-modal biometrics. Such limitations are noisy data, low distinctiveness, intra-user variation, non-universality of biometric characteristics, and vulnerability to spoof attacks.

Information fusion is used to produce a unified biometric decision based on multiple biometric sources. Simple approaches such as the sum rule score-level fusion proved to achieve high performance compared to more sophisticated approaches [1]. A step ahead is the weighted sum rule where each biometric source is weighted to indicate its relative importance, and thus contribution, to the final fused biometric decision.

Searching for the optimal weights combination can be done by exhaustive search to find optimal solution on training data. However, as will be shown in the next sections, this solution has low generalization ability. Weighting methods based on the statistics of the imposter and genuine scores distributions showed better and more generalized performance. Weighting based on the equal error rate of biometric sources is widely used [2] along with approaches based on D-Prime calculations [3] and Fisher discriminant ratio [4].

A comparative study by Chia et al. [5] discussed the performance of the most common weighting approaches and proposed a weighting algorithm that depends on the width of the overlapped area between the imposter and genuine scores distributions. Other works proposed fusion approaches based on non-linear combiners [6]. Benchmarking quality-based multi-biometric fusion was also discussed by Poh et al. [7].

In this work, a number of previously proposed weighting approaches are discussed and evaluated. Two new weighting techniques are presented and evaluated on two multi-biometric score databases and compared to the existing approaches. In the next Section 2, previously proposed baseline weighting algorithms are discussed along with the proposed approaches. Section 3 presents the performed experiments and the achieved results. Finally, a conclusion of this work is drawn.

2. APPROACH

A score-level multi-biometric decision is produced by combining the scores produced by different biometric sources. Those biometric sources vary in performance and thus should have different effect on the fused decision. Within the widely used linear combination fusion, each biometric source k is assigned a weight w_k that represents its relative effect in the final biometric decision. This section discusses solutions for biometric source weights calculations.

2.1. Baseline Weighting Approaches

In the following, a list of biometric source weighting approaches is presented. Those approaches present the state of

The work leading to these results has received funding from the European Community's Framework Programme (FP7/2007-2013) under grant agreement n° 261712 for the collaborative project CAPER.

the art and common practices in multi-biometric fusion using the weighted sum rule.

a) EER weighted (EERW): equal error rate is the common value of the false acceptance rate (FAR) and the false rejection rate (FRR) at the operational point where both FAR and FRR are equal. EER weighting was used to linearly combine biometric scores in the work of Jain et al. [2]. The EER is inverse proportional to the performance of the biometric source. Therefore, for a multi-biometric system that combines N biometric source, the EER weight for a biometric source k is given by:

$$w_k = \frac{\frac{1}{EER_k}}{\sum_{k=1}^{N} \frac{1}{EER_k}} \tag{1}$$

b) D-Prime weighted (DPW): D-Prime is used to measure the separation between the genuine and the imposter scores [3]. High separation indicates a higher performance of the biometric source. Given that σ_k^G and σ_k^I are the genuine scores and imposter scores standard deviations and μ_k^G and μ_k^I are their mean values, the D-prime is given by:

$$d'_k = \frac{\mu_k^G - \mu_k^I}{\sqrt{(\sigma_k^G)^2 + (\sigma_k^I)^2}} \tag{2}$$

and it is directly proportional to the performance of the biometric source and thus the weight can be calculated as:

$$w_k = \frac{d'_k}{\sum_{k=1}^{N} d'_k} \tag{3}$$

c) NCW weighted (NCWW): the Non-Confidence Width Weight was proposed by Chia et al. [5] to weight biometric sources for score-level multi-biometric fusion. NCW corresponds to the width of the overlap area between the genuine and imposter scores distributions. Given that Max_k^I is the maximum imposter score and Min_k^G is the minimum genuine score, NCW is given by:

$$NCW_k = Max_k^I - Min_k^G \tag{4}$$

as the NCW is inverse proportional to the biometric source performance, the weights based on the NCW is given as:

$$w_k = \frac{\frac{1}{NCW_k}}{\sum_{k=1}^{N} \frac{1}{NCW_k}} \tag{5}$$

d) FDR weighted (FDRW): the Fisher Discriminant Ratio as described by Lorena and Carvalho [8] and used by Poh et al. [4] measures the separability of classes, here genuine and imposter scores. The higher the separability, the higher is the biometric source performance. The FDR and the corresponding weights are given as:

$$FDR_k = \frac{(\mu_k^G - \mu_k^I)^2}{(\sigma_k^G)^2 + (\sigma_k^I)^2} \tag{6}$$

$$w_k = \frac{FDR_k}{\sum_{k=1}^{N} FDR_k} \tag{7}$$

e) Brute force weighted (BFW): here, the weights are assigned by brute force search for optimal weights (weights that produces lowest EER) on the training data. This method is computationally expensive especially for higher order multi-biometrics, therefore, only bi-modal biometric fusion were evaluated by BFW in this work.

f) Equal weighted (EQW): equal weighting assigns equal weights to all biometric sources under the assumption that all sources have the same contribution to the final fused biometric decision. This is usually used when no sufficient information (data) are available for the biometric sources in hand.

The use of EER, D-Prime and FDR is common practice for weighting biometric sources. However, more recent approaches just as the NCW proved superiority over such approaches [5]. Using brute force to assign weights has high computational expense and produce less generalized results as shown later in Section 3. The high performance of NCWW is however fragile as the NCW calculation depends on extrema values of comparison scores, which makes its performance very sensitive to outliers in training data.

2.2. Proposed Weighting Approaches

In this work, two weighting algorithms are proposed based on the properties of the genuine and imposter comparison scores distributions. First is the Mean-to-Extrema weighting (MEW) that depends on the mean values of the distributions with respect to their extremas. The second is the Overlap Deviation weighting (OLDW) that tries to avoid depending on unstable information such as distribution extrema (e.g. NCWW), and rather depends on more robust measures.

g) Mean-to-Extrema weighted (MEW): based on the assumption that a biometric source with low performance produces genuine score distribution that has a wide mean-to-min ranges and a wide mean-to-max imposter scores distribution range. The genuine mean-to-min range represents the difference between the mean of the genuine scores distribution and the minimum value (least correct) of the distribution. The same applies for the mean-to-max range in the imposter scores distribution.

The MEW is based on the width of the area between the mean of the imposter scores distribution and its maxima. It also considers the width of the area between the mean of the genuine scores distribution and its minima. This aims at focusing on the overlap area and its neighbor in both distributions. The MEW is formulated as:

$$ME_k = (Max_k^I - \mu_k^I) + (\mu_k^G - Min_k^G) \tag{8}$$

$$w_k = \frac{\frac{1}{ME_k}}{\sum_{k=1}^{N} \frac{1}{ME_k}} \tag{9}$$

h) Overlap Deviation weighted (OLDW): this weighting approach is based on two assumptions, first is the inverse relation between the performance of a biometric system and the standard deviation of the overlap are in its genuine-imposter scores distributions. The second assumption is the inverse relation between the EER value produced by a certain biometric source and its performance.

Overlap deviation tries to capture the properties of the overlap area between the imposter and genuine scores distributions without depending on singular extrema values. It also integrates the overall performance (FRR and FAR as an EER value) of the biometric source. Taking the standard deviation of this area aims at reducing the sensitivity to outliers in the data with respect to considering the width of the area. Including the overall verification performance of the biometric source (EER) in the weighting process aims at creating a better generalized solution.

Given the imposter scores S_k^I, the genuine scores S_k^G, the equal error rate EER and the score threshold at the equal error operating point T, the OLDW can be given as:

$$OLD_k = \sigma(\{S_k^I \mid S \geq T\} \cup \{S_k^G \mid S < T\}) \times EER \quad (10)$$

$$w_k = \frac{\frac{1}{OLD_k}}{\sum_{k=1}^{N} \frac{1}{OLD_k}} \quad (11)$$

In the next Section 3, the experiment design is introduced. The performance achieved by the different weighting algorithms discussed will be presented.

3. EXPERIMENT AND RESULTS

The proposed approaches for multi-biometric source weighting are general and can be applied to any number of biometric sources. However, the presented results focus on the case of bi-modal biometrics to investigate the performance away from high order complexities. Moreover, the performance of high order multi-biometric scenarios is also investigated.

Two multi-biometric scores databases were used to develop and evaluate the discussed solutions in order to assist the generalization capabilities of those solutions.

The first database is the Extended Multi Modal Verification for Teleservices and Security applications database (XM2VTS) [9] [10]. The Lausanne Protocols I (LP1) partition of the XM2VTS database was used in the experiment. This partition contains comparison scores produced by five face (F0 - F4) and three speech (S5 - S7) baseline experts. The evaluation and development sets defined by the authors were used in the performed experiments. The experiments here considered all possible pairs between face and speech matchers as well as the fusion of all matchers. The database contains 295 individuals, which results in 1000 genuine and $151,800$ imposter scores. For more details about the XM2VTS score database, one can refer to the work of Poh et al. [9].

The second database used is the Biometric Scores Set BSSR1 - multimodal database [11]. The database contains comparison scores for left and right fingerprints (Fli and Fri) and two face matchers (Fc and Fg). BSSR1 - multimodal database contains 517 genuine and $266,772$ impostor scores. The experiments here considered all possible pairs between finger and face matchers as well as the fusion of all matchers. To evaluate the statistical performance of the proposed solutions, the database was splitted into three equal-sized partitions. Experiments were performed on all possible fold combinations were one partition is used as evaluation set and the other two are used as a development set. The reported results are the averaged results of the three evaluation/development combinations.

Min-max normalization was used to bring comparison scores produced by different biometric sources to a comparable range [12]. Min-max normalized score S' is given as:

$$S' = \frac{S - min\{S_k\}}{max\{S_k\} - min\{S_k\}} \quad (12)$$

Where $min\{S_k\}$ and $max\{S_k\}$ are the minimum and maximum value of scores existing in the training data of the corresponding biometric source.

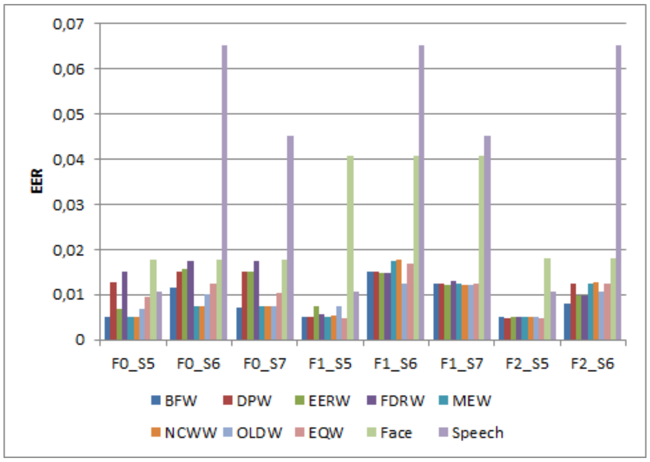
To produce the fused scores, the weighted sum rule (linear combination) was used. The weighted sum rule assigns each score value S_k with the weight of its source w_k. The weights w_k are calculated from the training data of each biometric source as discussed in Section 2. The fused score by the weighted sum rule F for N score sources is given as:

$$F = \sum_{k=1}^{N} w_k S_k, k = \{1, \ldots, N\} \quad (13)$$

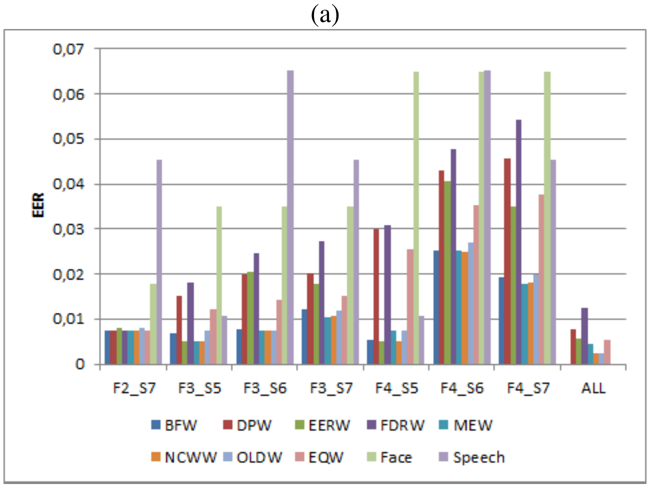
The performance of the fusion process under different weighting approaches is evaluated under verification scenario and presented as EER values and as Receiver Operating Characteristic (ROC) curves.

For each of the databases, all bi-modal combinations are evaluated along with the overall fusion of all available sources. As expected, the results show the advantage of multi-biometrics on the verification performance.

The EER values obtained from the XM2VTS database are shown in Figure 1. One can notice the high performance of NCWW and OLDW both scoring 0.25% EER for the overall fusion evaluation followed by the MEW with 0.46% EER and far from the commonly used DPW with 0.75% EER. In the bi-modal evaluation, NCWW showed high performance in many combinations closely followed by a stable performance by the OLDW. It must be noticed that in some scenarios such as in $F1\_S6$ and $F2\_S6$ NCWW performed worse than most approaches while OLDW sustained stable high performance.



(a)



(b)

Fig. 1. Equal error rates achieved on the XM2VTS database: The rates shown here represent all possible bi-modal combinations of face matchers (F0 - F4) and speech matchers (S5 - S7) in the XM2VTS database and the results achieved by the fusion of all eight available sources.

EER values obtained from different approaches using the BSSR1 database are shown in Figure 2. The figure shows the superiority of the OLDW approach in most cases with stable performance compared to the NCWW approach. In the overall fusion evaluation, the OLDW score the best performance at 0.21% EER followed by EERW, EQW and MEW while the NCWW score 0.37% EER. The fluctuation in the NCWW performance, with respect to that of the OLDW, may be related to its dependence on extrema values that are more vulnerable to outliers than the measures used to calculate the OLDW.

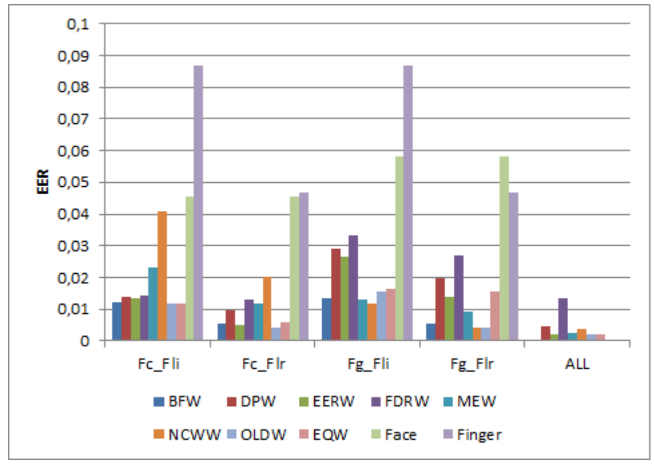


Fig. 2. Equal error rates achieved on the BSSR1 database: The rates shown here are for all possible bi-modal combinations of face matchers (Fc and Fg) and finger matchers (Fli - Flr) in the BSSR1 database and the results achieved by the fusion of all four available sources.

Results of evaluation over the XM2VTS database using different weighting approaches to fuse all available sources (five face and three speech) are also shown as ROC curves (Figure 3) to investigate the performance under different operational points. One very low false acceptance rates (FAR), the proposed OLDW performs the best. While the FAR values get higher, the lowest false rejection rate (FRR) is achieved by the NCWW and the proposed OLDW and MEW approaches.

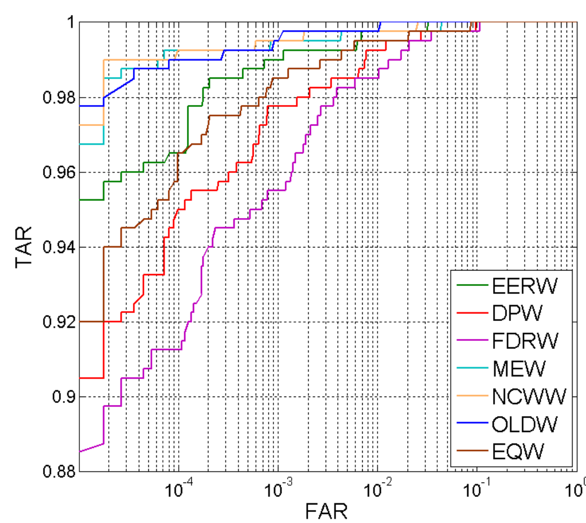


Fig. 3. ROC curves achieved on the XM2VTS database: The curves shown here represent the performance of the fusion of all eight (five face and three speech) available sources using different weighting approaches.

The ROC curves achieved on the BSSR1 database are shown in Figure 4. Those curves are graphically averaged curves over the three testing folds of the database in a similar manner to vertical averaging discussed in [13]. One can notice the superiority of the OLDW performance, especially at low FAR, followed by the MEW and the NCWW.

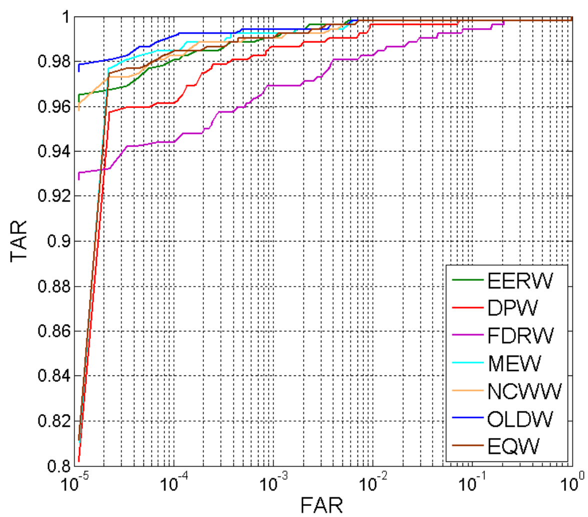


Fig. 4. ROC curves achieved on the BSSR1 database: The curves shown here represent the performance of the fusion of all four available sources (two face matchers and two fingerprint matchers) using different weighting approaches.

4. CONCLUSION

This work presented a weighting approach for biometric sources within a weighted sum rule multi-biometric fusion solution. The proposed solution aimed at being more robust to outliers in the data and having higher generalization capabilities compared to baseline approaches. To achieve this, the approach considered the standard deviation of the overlapped area between genuine and imposter scores distributions. It also considered the overall performance (as EER) achieved by each biometric source. Tests were carried on two different multi-biometric score databases and the results were presented as EER values and ROC curves. The results showed a satisfying and robust performance compared to the state of the art approaches.

REFERENCES

[1] Arun Ross and Anil K. Jain, "Information fusion in biometrics," *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2115–2125, 2003.

[2] Anil Jain, Karthik Nandakumar, and Arun Ross, "Score normalization in multimodal biometric systems," *Pattern Recognition*, vol. 38, no. 12, 2005.

[3] Robert Snelick, U. Uludag, Alan Mink, M. Indovina, and A. Jain, "Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 3, 2005.

[4] Norman Poh and Samy Bengio, "A study of the effects of score normalisation prior to fusion in biometric authentication tasks," Idiap-RR Idiap-RR-69-2004, IDIAP, 0 2004.

[5] Chaw Chia, N. Sherkat, and L. Nolle, "Towards a best linear combination for multimodal biometric fusion," in *Pattern Recognition (ICPR), 2010 20th International Conference on*, 2010, pp. 1176–1179.

[6] Norman Poh and Samy Bengio, "Non-linear variance reduction techniques in biometric authentication," in *In Workshop on Multimodal User Authentication (MMUA 2003*, 2003, pp. 123–130.

[7] N. Poh, T. Bourlai, J. Kittler, L. Allano, F. Alonso-Fernandez, O. Ambekar, J. Baker, B. Dorizzi, O. Fatukasi, J. Fierrez, H. Ganster, J. Ortega-Garcia, D. Maurer, A.A. Salah, T. Scheidat, and C. Vielhauer, "Benchmarking quality-dependent and cost-sensitive score-level multimodal biometric fusion algorithms," *Information Forensics and Security, IEEE Transactions on*, vol. 4, no. 4, pp. 849–866, Dec 2009.

[8] Ana Carolina Lorena and André C. P. L. F. de Carvalho, "Building binary-tree-based multiclass classifiers using separability measures," *Neurocomput.*, vol. 73, no. 16-18, pp. 2837–2845, Oct. 2010.

[9] Norman Poh and Samy Bengio, "Database, protocols and tools for evaluating score-level fusion algorithms in biometric authentication," *Pattern Recognition*, vol. 39, no. 2, pp. 223 – 233, 2006.

[10] Norman Poh, Thirimachos Bourlai, and Josef Kittler, "A multimodal biometric test bed for quality-dependent, cost-sensitive and client-specific score-level fusion algorithms," *Pattern Recogn.*, vol. 43, no. 3, pp. 1094–1105, Mar. 2010.

[11] *National Institute of Standards and Technology: NIST Biometric Scores Set.*

[12] Naser Damer, Alexander Opel, and Alexander Nouak, "Performance anchored score normalization for multi-biometric fusion," in *Advances in Visual Computing*, pp. 68–75. Springer Berlin Heidelberg, 2013.

[13] Norman Poh, Alvin Martin, and Samy Bengio, "Performance generalization in biometric authentication using joint user-specific and sample bootstraps," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 3, pp. 492–498, 2007.