Multibiometrics enhancement using quality measurement in score level fusion

Saliha Artabaz[1], Layth Sliman[2], Hachemi Nabil Dellys[1], Karima Benatchba[1], Mouloud Koudil[1].

[1] Ecole nationale Supérieure d'Informatique ESI, Oued-Smar, Alger, Algérie.
[2] Ecole d'ingénieur généraliste en informatique et technologies du numérique Efrei
30-32 avenue de la République, 94800, Villejuif, Paris, France.
[1]{s_artabaz, k_benatchba, h_dellys, m_koudil}@esi.dz
[2]layth.sliman@efrei.fr

Abstract. Multi-biometrics is a solution which is sought to overcome functional and security deficiency in a baseline biometric configuration. In this paper, we propose a multi-biometrics scheme and we apply the cross validation between two databases to study the Equal Error Rate improvement of score level fusion. Our fusion function is constructed using an evolutionary GA on the XM2VTS score database. The best one is tested on a sub-sequence of the BioSecure Score database. As this database offers quality measurement, we transform our function into a weighted function with user-specific approach to study performance enhancement with quality integration. The results are significantly improved with a high confidence and quality measurement becomes inherent to reduce recognition errors.

Keywords: Weighted fusion, user-specific, tree structure, cross-validation.

1 Introduction

Multi-biometrics is well-considered to deal with traditional biometric systems drawbacks. Major motivations are performance improvement and universality fulfilment. In fact, multiple biometrics allows handling enrolment errors by offering alternative possibilities when a biometric is missing. In addition, experiments show that combining different evidences can enhance performances [1-6]. However, the fused evidences cannot be equivalent and must not be considered as the same in the fusion process [8]. That is why recent researches focus on evaluating quality of each sub-system as entry of the fusion process. As the quality is one of the main factors affecting the overall performance of biometric systems, measuring quality information aims to overcome fusion process errors. Multi-biometrics fusion considers different levels: image, feature, score, decision and rank level. The most used in the literature is the score level [16, 17]. At this level, combined scores are easier to fuse and provide rich information at the same time [4]. Furthermore, the ease of accessing to scores and good performances that outperform other levels make it as the best level [17]. The quality assessment considers three points of view depending on the considered information:

adfa, p. 1, 2011.

features, relation between sample and source and impact on the performances. In this paper, we are interested in the score fusion with quality weighting. We conduct a previous work in fusion level presented in [7] by optimizing a fusion function using a tree structure. This work gives sufficient results on the XM2VTS[1] score database. An extended evaluation is needed to prove the effectiveness of the proposed approach. In addition, we aim at optimizing the score distribution by integrating the template-query[2] quality as weights. We choose the Biosecure score database [13] as it is, to the best of our knowledge, the only database that offers quality and cost evaluation. We propose a weighted function based on quality measurement between sample and source. This is simple to apply in real case by including the quality value in the template and integrating the measurement function in the biometric system. The proposed approach outperforms the basic biometric sub-systems.

This paper is organized as follows: First, in section 2, we introduce the studied field. Section 3 illustrates the proposed fusion approach and explains database used to conduct experiments. After that, we describe experimental results in section 4. Finally, we conclude and list some perspectives of our work.

2 Multi-biometrics and quality fusion

Multi-biometrics is an emerging field that deals with unimodal biometric system problems. Many researchers conduct their experiments on different systems to demonstrate the effectiveness of the fusion in performances enhancement. However, their approaches combine heterogeneous systems and data. The fusion requires normalized data to avoid biased results. Furthermore, weighting systems outputs is required to preserve and improve resulting system performances. This can be done using statistical estimation from system performances or quality measurement [3,5]. The quality measurement [14] is a promising key that can be used as predictor of the system performances. The authors in [14] define sample quality as "*scalar quantity that is related monotonically to the performance of biometric matchers*". They discuss the effectiveness of a quality evaluation and the different ways to provide this scalar. The sample quality is a result of different points of view [8]:

- Character: refers to the quality (properties) of the physical features of the individual.
- Fidelity: refers to the degree of similarity between a biometric sample and its source (faithfulness).
- Utility: refers to the impact of the individual biometric sample on the overall performance of a biometric system (predicted contribution in performance).

The usefulness of quality measurement and integration is unavoidable in the fusion process since fused systems are heterogeneous. The computed scalars for each sample can be used as weights to fix unbalanced fusion function. This is the subject of many

---

1 XM2VTS is score database built with Lausanne Protocol and based on Face and speech [18].

2 Template is the data sample used to represent the claimed identity and Query is the sample obtained from the <true ID>.

recent researches in order to improve fusion performances [9]. The authors in [9] explain the main raisons that led to use quality measures. They focus on the need to design systems that work in an unsupervised environment, to match with new biometric systems requirements such as portable or low-cost devices, remote access, distant acquisition or forensics. [10] Uses quality information to switch between different system modules depending on the data source. In this work, the authors use linear logistic regression to obtain an efficient combination of matching scores. In [11], authors propose weighting modalities with ancillary information based on relative degradation to enhance performances under noise conditions. [12] processes the data uncertainty concept and fusion confidence with uncertainty function of factors affecting system performance. Their proposition is built upon Dempster-Shafer approach. [15] Uses Score Reliability Based Weighting (SRBW) to estimate matcher weights. The reliability is computed using raw matching scores of each system.

3 Proposed scheme

Fusion function is the major step to define multi-biometric system. The rule-based function is considered as the best and the most intuitive method to use, exceptionally the sum and weighted sum. In this paper, we aim at testing fusion function based on tree structure with advanced experiments [7]. The function is computed using evolutionary algorithm 'GA'. We use tree structure to generate and apply modifications on fusion function with mutation and crossover operations. Our preliminary tests give high performances with 50 generations. In practice, the fitness function stabilizes within this number. We conduct additional tests with extended number of generations up to 100. We adjust the convergence criteria to 0,001.

Fig. 1 illustrates initial population and its evolution with GA simulation according to the crossover operation used to evolve the population, and roulette selection upplied to get the best population list with error under 0,05. A ratio of this population number to the total number of individual is computed. The experimented process progresses to reach a steady range between 60% and 100%. The graph oscillations show the progress of the GA algorithm, which does not converge systematically.

This simulation allows a significant improvement of the fused system. Fig. 2 shows the min, max, mean and standard deviations. The max and standard deviation curves illustrate better results and stabilize within 100 generations. We reach an HTER=0,0018 with 60% improvement (Last HTER=0,0045).

We select the best tree obtained from the simulation to establish a cross-validation on another database in order to prove the efficiency of the proposed method. In our work, we aim at studying the impact of quality integration on multi-biometric system. We test the use of quality measurement for weighting our fusion function. We use a user-specific weighting in order to improve baseline systems performances. Which means that weights are adapted to quality measurement of each individual. Each computed weight reflects the sample fidelity to the claimed ID. It represents the gap between sample and its source. We compute the weight as follows:

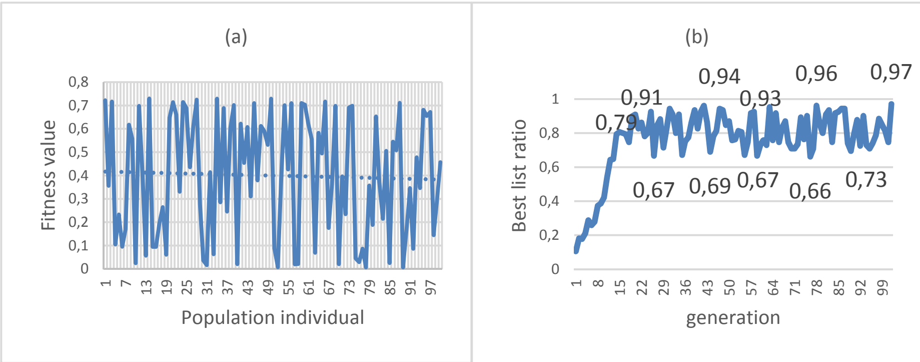

Fig. 1. (a) Fitness of initial population (b) Evolution of best population ratio during generations.

$$\left(\frac{1}{2} * \left((x_t-\bar{x})^2 + \left(x_q-\bar{x}\right)^2\right)^{1/2}\right)^{-1} \tag{1}$$

where x_t is the template quality (claimed id) and x_q is the query quality (true id).

We use this equation to compute inverse of the standard deviation between the template quality and query quality. A small value means that the two measurements are very different and vice-versa.

To normalize the quality, we use the standard deviation between max and min values computed from the development set. The resulting score fusion is applied on the normalized scores, with normalization to 1 value, using the following function:

$$avg(S_1, avg(\min(\min(S_2,S_3) + S_4 * S_5, S_6), \min(S_7,S_8))) \tag{2}$$

where S_i are the selected scores.

4 Experiments and discussion

We choose the Biosecure score database [13] as it is the only score database that offers a quality evaluation of its sets. We perform the experiments, using the Biosecure

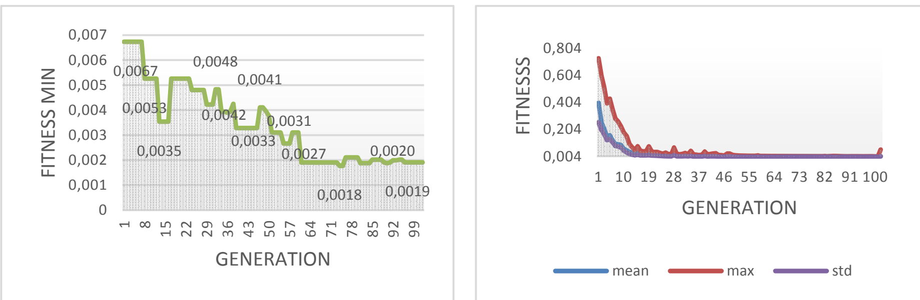

Fig. 2. Statistical values of obtained population during the 100 generations.

protocol [13] on two sessions of development and evaluation set, according to these steps:

1. Generate 1000 different configurations to select 8 scores from 24 scores of the database;
2. Test the baseline systems of these configurations;
3. Test each configuration with the proposed fusion function on three multi-biometric systems using development set (session 1) to fix experimental parameters:
   (a) fusion of the baseline scores;
   (b) fusion of the normalized scores;
   (c) fusion using the weighted function with quality scalars with normalized and non-normalized scores.
4. Check results on session 2 to avoid biased performances;
5. Repeat tests (a-c) on evaluation set to render system performances (session 1 and 2);
6. Replace dummy values in the normalized baseline system with max and min value according to their sign.

The preliminary results seem interesting since it optimizes scores under the mean of scores 0,54 and 0,58 for session 1 and session 2 respectively. The normalization involves 60 % improvement. Results for session 2 are quite better with (AVG=8%, STD=22%) vs. (AVG=16%, STD=26%) for session 1 (see Fig. 3).
The normalization of baseline systems before fusion reduce fused system errors. However, the curve is not regular. As shown in Fig. 3, the curve for session 2 is chunked into two parts: best EER and worst EER. This is due to the standard deviation that exceeds STD in quality based system. By applying quality measurement, we achieve significant improvement. Despite the worst values above the baseline values, the curve is regular around the mean, which is equal to 9% and 12% for test session 1 and session 2 respectively. This can be seen by the small standard deviation (STD=11%) comparing to normalization results. These results outperform the fusion of the best-selected systems, which achieve 6% for EER. The quality weighting reaches an EER equal to 0,6%. In addition, our method succeed to enhance worst systems (82% of generations) to achieve an EER between 0 and 20%.

As it is recommended to use session 2 of the evaluation set, we apply the proposed fusion on it to get performances (EER and ROC curve). Fig. 4 shows a comparison

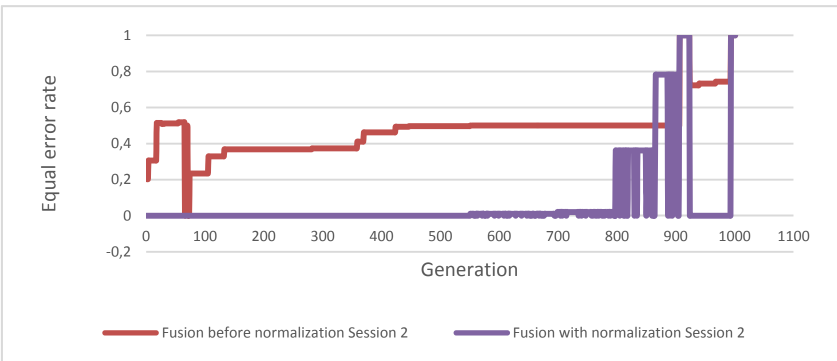

Fig. 3. Normalization impact on EER for 1000 configurations with development set (session 2).

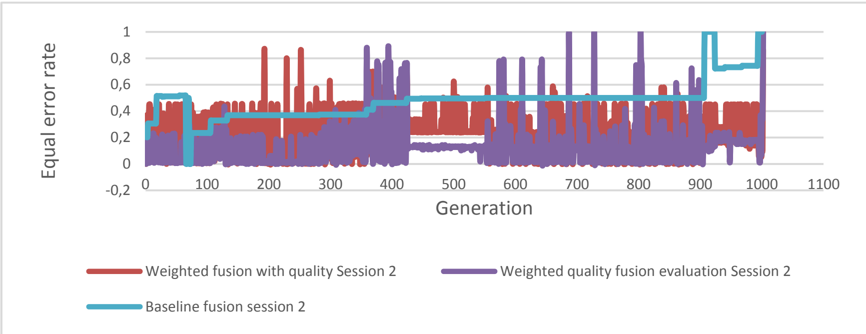

Fig. 4. Weighted fusion function applied on development and test set for session 2.

between development set and evaluation set (session 2). The two curves seem to be similar with a comparable standard deviation (16 for the first and 15 for the last). However, the mean decreases for the evaluation set to half value (12 to 23). We can notice that system performances are preserved even tested on different sets.

The ROC curve, of one achieved fusion system under the mean EER=1,3%, is shown in Fig. 5. It compares between the proposed approach and the eight selected baseline systems. The weighing achieve an important enhancement compared to initial systems. The AUC is much higher for these systems than the fused weighted system. This illustrates the importance of the quality information in the fusion process.

5 Conclusion and future work

In this paper, we propose a weighted fusion function based on primitive fusion rule. We use Genetic algorithm to explore search space and find out the best rules combination using the XM2VTS. Next, we apply user-specific weighting on the acquired function. Experiments on the Biosecure score database provide promising improvement. In

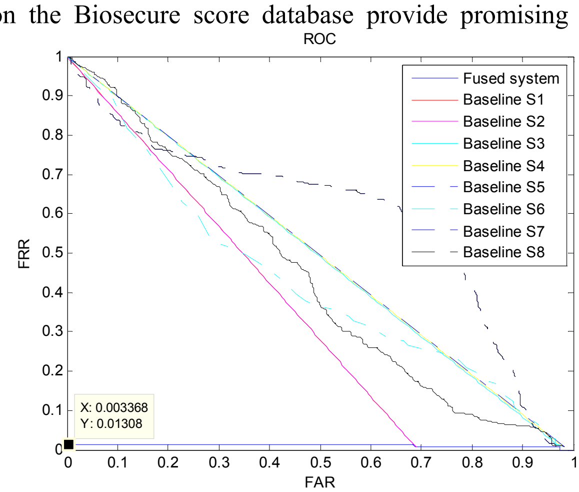

Fig. 5. ROC curves (EER=0,013) for weighted fusion function and selected baseline systems.

addition of multi-biometric advantages, these results reach a significant optimization of EER. This work can be extended to compare with other normalization methods and study the impact on density distribution. We can select a sub-list of the obtained trees in order to establish comparative study between them.

References

1. AlMahafzah, H. & AlRawashdeh, M. Z. Performance of Multimodal Biometric Systems at Score Level Fusion *Wireless Communications, Networking and Applications: Proceedings of WCNA 2014, Springer India,* **2016**, 5, 903-913
2. Ghulam Mohi-ud-Din, S., Mansoor, A.B., Masood, H. Personal identification using feature and score level fusion of palm- and fingerprints *Signal, Image and Video Processing, Springer,* **2011**, 477–483
3. Anzar, S. M. & Sathidevi, P. S. Optimization of Integration Weights for a Multibiometric System with Score Level Fusion *Advances in Computing and Information Technology: Proceedings of the Second International Conference on Advances in Computing and Information Technology (ACITY) July 13-15, 2012, Chennai, India - Volume 2, Springer Berlin Heidelberg,* **2013**, 833-842
4. Conti, V.; Militello, C.; Sorbello, F. & Vitabile, S. A Frequency-based Approach for Features Fusion in Fingerprint and Iris Multimodal Biometric Identification Systems *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews),* **2010**, *40*, 384-395
5. Damer, N. & Opel, Multi-biometric Score-Level Fusion and the Integration of the Neighbors Distance Ratio *Image Analysis and Recognition: 11th International Conference, ICIAR 2014, Vilamoura, Portugal, October 22-24, 2014, Proceedings, Part II, Springer International Publishing,* **2014**, 85-93
6. Eskandari, M. & Toygar, Ö. Score Level Fusion for Face-Iris Multimodal Biometric System Information Sciences and Systems 2013: Proceedings of the 28th International Symposium on Computer and Information Sciences, Springer International Publishing, **2013**, 199-208
7. Artabaz, S.; Sliman, L.; Benatchba, K.; Dellys, H. N. & Koudil, M. Score Level Fusion Scheme in Hybrid Multibiometric System *Advances in Visual Informatics: 4th International Visual Informatics Conference, IVIC 2015, Bangi, Malaysia, November 17-19, 2015, Proceedings, Springer International Publishing,* **2015**, 166-177
8. El-Abed, M.; Charrier, C. & Rosenberger, C. Quality assessment of image-based biometric information *EURASIP Journal on Image and Video Processing,* **2015**, *2015*, 1-15
9. Alonso-Fernandez, F.; Fierrez, J. & Bigun, J. Quality Measures in Biometric Systems *Encyclopedia of Biometrics, Springer US,* **2015**, 1287-1297
10. Alonso-Fernandez, F.; Fierrez, J.; Ramos, D. & Gonzalez-Rodriguez, J. Quality-Based Conditional Processing in Multi-Biometrics: Application to Sensor Interoperability *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans,* **2010**, *40*, 1168-1179
11. Mohammed Anzar, S. T. & Sathidevi, P. S. On combining multi-normalization and ancillary measures for the optimal score level fusion of fingerprint and voice biometrics *EURASIP Journal on Advances in Signal Processing,* **2014**, 1-17
12. Nguyen, K.; Denman, S.; Sridharan, S. & Fookes, C. Score-Level Multibiometric Fusion Based on Dempster Shafer Theory Incorporating Uncertainty Factors *IEEE Transactions on Human-Machine Systems,* **2015**, *45*, 132-140

13. N. Poh, T. Bourlai, and J. Kittler, “A Multimodal Biometric Test Bed for Quality-dependent, Cost-sensitive and Client-specific Score-level Fusion Algo-rithms,” in Pattern Recognition Journal, accepted, 2009
14. P. Grother et E. Tabassi, Performance of Biometric Quality Measures, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **2007**, 29, 531-543
15. Kabir, W.; Ahmad, M. O. & Swamy, M. N. S. Score reliability based weighting technique for score-level fusion in multi-biometric systems *2016 IEEE Winter Conference on Applications of Computer Vision (WACV),* **2016**, 1-7
16. Lip, C. C. & Ramli, D. A. Comparative Study on Feature, Score and Decision Level Fusion Schemes for Robust Multibiometric Systems *Frontiers in Computer Education, Springer Berlin Heidelberg,* **2012**, 941-948
17. A. K. Jain and A. Ross, "Introduction to biometrics" in Handbook of Biometrics, pp. 91-142, 2008, Springer-Verlag.
18. Norman Poh and Samy Bengio, Database, Protocol and Tools for Evaluating Score-Level Fusion Algorithms in Biometric Authentication, Pattern Recognition, Volume 39, Issue 2, Pages 223-233, 2006.