

Experimental Evaluation of Matching-Score Normalization Techniques on Different Multimodal Biometric Systems

Slobodan Ribaric and Ivan Fratric
 Faculty of Electrical Engineering and Computing
 University of Zagreb
 Zagreb, Croatia
 slobodan.ribaric@fer.hr, ivan.fratric@fer.hr

Abstract—The goal of this work is experimental evaluation of matching-score normalization techniques for the following three multimodal biometric systems: a verification system based on the fusion of face and palmprint principal line features, an identification system based on eigenfingers and finger-geometry, and an identification system based on eigenpalm and eigenfinger features. The following normalization techniques are tested: Bayes-based normalization and several heuristic techniques (min-max, z-score, median-MAD, double-sigmoid, tanh, and piecewise-linear). The results of evaluation are represented by means of system performance (expressed by ROC, EER and minTER) and F-statistics.

I. INTRODUCTION

A multimodal biometric system [1] requires an integration scheme to fuse the information obtained from the individual modalities. There are various levels of fusion [2, 3] - from sensor to decision level. Fusion at the matching-score level is the most popular and frequently used method because of its good performance, intuitiveness and simplicity [4, 5, 6].

When using the method of fusion at the matching-score level a normalization step is generally required for the following reasons [5]: i) The matching scores at the output of the matchers for different modalities can be represented in different ways. For example, one matcher may output distances (as a measure of dissimilarity), while the others may output proximities (as a measure of similarity); ii) The matcher outputs can be in different numerical ranges and iii) The genuine and impostor matching scores from different modalities might not follow the same statistical distributions.

Fig. 1 shows the detail of multimodal system with fusion at the matching-score level. There is a set of M matchers, usually one for each modality. In general, the normalized score is obtained by using a normalization function. Let $r_i \in R_i$, $i = 1, 2, \dots, M$, where r_i is the output raw value of the matcher i . R_i is the set of all the raw output values of the corresponding matcher i (matching scores). Usually, $R_i \subset \mathfrak{R}^1$. The corresponding normalized

similarity matching scores are denoted as $n_i \in N_i$, where $N_i \subset \mathfrak{R}^1$ is the set of normalized matching scores of the matcher i . The normalization is defined as a function that maps R_i to N_i . The normalized matching scores n_i , $i = 1, 2, \dots, M$ are combined in the fusion module and the result of fusion is represented by the total similarity measure (TSM).

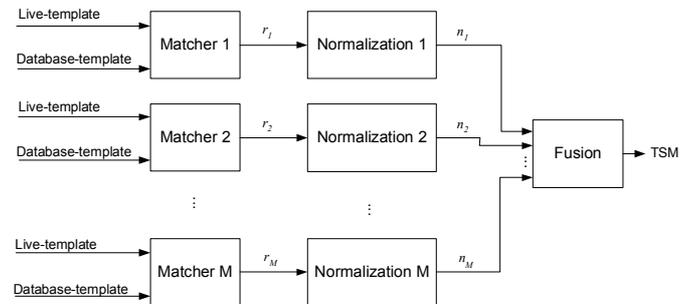


Figure 1. A detail of a multimodal system with fusion at the matching-score level

In our experiments the following normalization techniques are used: Bayes-based normalization [5] where the corresponding distributions are estimated based on the Parzen windows [7] and several heuristic normalization techniques [5, 6, 8] (min-max, z-score, median-MAD, double-sigmoid, tanh, and piecewise-linear). We tested these techniques on three prototypes of biometric systems that combine different modalities. We represented the experimental results by means of EER (Equal Error Rate), minTER (minimum Total Error Rate), ROC (Receiver Operating Characteristic) curves and F-statistics [9].

II. EXPERIMENTAL SETUP

The mentioned normalization techniques have been tested on the following multimodal biometric systems, which were developed in our laboratory:

- i) Verification system based on the fusion of palmprint principal lines and eigenface features [10] (we denoted this system by V-P-F in the remaining part of the paper). The

palmpoint matching-score is expressed as similarity measure, while the face matching-score is expressed as distance.

- ii) Identification system based on four eigenfingers and finger-geometry [11] (denoted by I-EF-FG). Eigenfingers are a novel features extracted by means of PCA from the strip-like area of the fingers [12]. All matching-scores are expressed as distances.
- iii) Identification system [12] based on five eigenfingers and eigenpalms (denoted by I-EF-EP). All matching-scores are expressed as distances.

The system i) was tested on the “chimerical” multimodal database created by pairing images from the XM2VTS face database [13] and our own palmpoint database (180 dpi grayscale images). 440 image pairs of 110 persons were used for training and 1048 image pairs of 131 persons were used for testing (there are 8 images per person). Five image pairs of each person were used for enrollment and the remaining three were used for testing.

The systems ii) and iii) were both tested on our own palmpoint database. 550 images of 110 persons were used for training, while 1926 images of 209 persons were used for testing. The test database was further divided into client and impostor parts. 127 persons with 10 images per person played the role of clients, while the remaining 82 persons with 8 images per person played the role of impostors. For the persons playing the role of clients, seven images were used for enrollment and the remaining three for testing. For the persons playing the role of impostors, all eight images were used for testing.

The motivation for testing the normalization techniques on the three multimodal systems is the attempt to investigate how the different normalization techniques affect the system performance for different modalities.

We represented the system performance by means of EER, minTER and ROC curves. EER (Equal Error Rate) is the error value for which the FAR (False Acceptance Rate) and FRR (False Rejection Rate) are equal. The minTER is the value of $FRR + FAR$, where this value is minimal. The ROC (Receiver Operating Characteristics) curves are graphs showing the dependency of GAR (Genuine Acceptance Rate, $GAR = 1 - FRR$) of the FAR.

We also use F-statistics to express the results of normalization. F-statistic [9] is a scalar measure used in speech recognition, information retrieval etc., to show the discriminatory power of some feature. It is defined as the ratio of between-class variance and the within-class variance for the given feature. In our experiments we calculated the F-statistics on the raw (unnormalized) scores for each modality, for normalized scores before the fusion, for the multidimensional distributions created by combining all matching scores, and finally, for the TSM (Total Similarity Measure) obtained after the fusion.

III. EXPERIMENTAL RESULTS

Figures 2, 3 and 4 represent ROC curves for different normalization techniques for the following systems: V-P-

F, I-EF-FG and I-EF-EP, respectively. Tables I, II and III show the corresponding values of EER and minTER for all three systems. The values of F-statistics for different modalities in three systems are given in Tables IV, V and VI.

As can be seen from the results, for the system V-P-F the Bayes-based normalization gives the best system performance (expressed by ROC, EER and minTER). Similarly, F-statistics gives the highest values for Bayes-based normalization.

In the remaining two systems (I-EF-FG and I-EF-EP) the heuristic normalization techniques give better system performance. In the system I-EF-FG, the best system performance is obtained using piecewise-linear normalization, while in the system I-EF-EP similar system performance is obtained using several heuristic normalization techniques. However, in these cases, the values of F-statistics do not follow the results of system performance and again give the highest values for Bayes-based normalization.

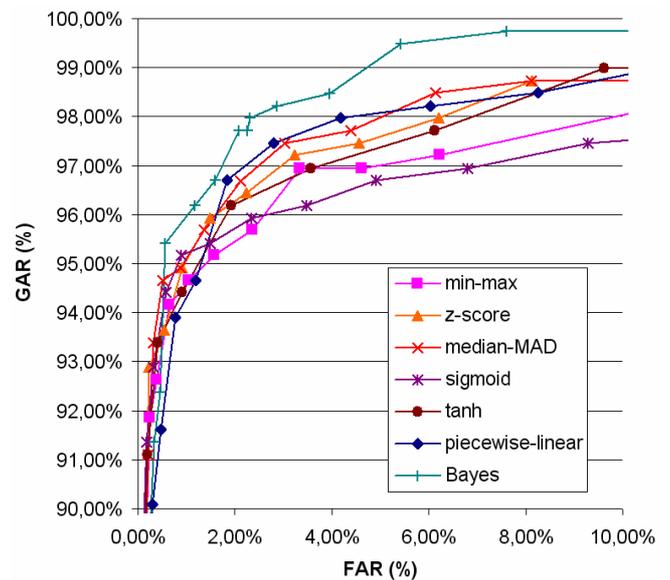


Figure 2. ROC curves for the verification system based on the fusion of face and palmpoint principal line features (V-P-F)

IV. CONCLUSION

We have experimentally evaluated seven matching-score normalization techniques for one multimodal verification system and two multimodal identification systems using different modalities. We represented the results of evaluation by means of system performance (expressed by ROC, EER and minTER) and F-statistics.

The experimental results suggest that there is no single normalization technique that would perform the best for all systems and for all modalities. Although Bayes-based normalization should give the best results in theory, in practice due to the difficulties in estimating the distributions, it does not always perform as well. F-statistics also suggest high discriminatory power of the Bayes-normalized scores but, in general, this does not guarantee the best system performance in term of error

TABLE I. EER AND MINTER FOR THE SYSTEM V-P-F

	piecewise-linear	min-max	median-MAD	double-sigmoid	tanh	zscore	Bayes
EER	2.79%	3.12%	2.79%	3.81%	3.05%	3.15%	2.29%
minTER	5.15%	6.39%	5.42%	5.72%	5.74%	5.56%	4.33%

TABLE II. EER AND MINTER FOR THE SYSTEM I-EF-FG

	piecewise-linear	min-max	median-MAD	sigmoid	tanh	zscore	Bayes
EER	1.05%	1.57%	1.57%	1.05%	1.31%	1.57%	1.38%
minTER	1.48%	2.26%	2.10%	1.75%	1.75%	2.10%	2.36%

TABLE III. EER AND MINTER THE SYSTEM I-EP-EF

	piecewise-linear	min-max	median-MAD	sigmoid	tanh	zscore	Bayes
EER	0.52%	0.52%	0.52%	0.79%	0.52%	0.52%	2.07%
minTER	0.70%	0.79%	0.78%	1.30%	0.70%	0.52%	3.72%

TABLE IV. F-STATISTICS FOR THE SYSTEM V-P-F

	before normalization	piecewise-linear	min-max	median-MAD	double-sigmoid	tanh	zscore	Bayes
Face	2.70	2.73	2.70	2.70	4.04	2.70	2.70	3.45
Palm	7.51	8.22	7.51	7.51	4.33	7.51	7.51	10.86
Face+Palm (2D)	2.70	5.41	4.97	5.98	4.18	4.78	5.02	10.86
TSM	N/A	10.70	9.41	10.12	8.39	9.20	9.46	10.86

TABLE V. F-STATISTICS FOR THE SYSTEM I-EF-FG

	before normalization	piecewise-linear	min-max	median-MAD	sigmoid	tanh	zscore	Bayes
little finger	3.05	3.65	3.05	3.05	3.38	3.13	3.05	3.91
ring finger	3.82	4.49	3.82	3.82	4.16	3.99	3.82	5.91
middle finger	4.70	5.83	4.70	4.70	4.56	4.86	4.70	6.34
index finger	3.46	3.88	3.46	3.46	3.58	3.59	3.46	3.49
geometry	2.80	6.71	2.80	2.80	5.72	2.86	2.80	29.80
4 fingers+geometry (5D)	3.45	4.83	3.59	3.58	4.31	3.08	3.63	6.33
TSM	8.05	11.90	7.82	8.12	10.70	8.03	7.83	16.01

TABLE VI. F-STATISTICS FOR THE SYSTEM I-EP-EF

	before normalization	piecewise-linear	min-max	median-MAD	sigmoid	tanh	zscore	Bayes
little finger	3.05	3.65	3.05	3.05	3.38	3.13	3.05	3.91
ring finger	3.82	4.49	3.82	3.82	4.16	3.99	3.82	5.91
middle finger	4.70	5.83	4.70	4.70	4.56	4.86	4.70	6.34
index finger	3.46	3.88	3.46	3.46	3.58	3.59	3.46	3.49
thumb	1.81	2.05	1.81	1.81	1.52	1.68	1.81	2.01
palm	4.83	5.80	4.83	4.83	6.17	5.14	4.83	4.47
5 fingers+palm (6D)	3.78	4.04	3.52	3.62	4.02	3.67	3.59	4.11
TSM	8.21	10.45	7.95	8.01	9.56	8.26	7.97	12.17

rates. In order to build a high-performing multimodal biometric system, different normalization techniques should be considered and evaluated in order to select the best-performing one.

REFERENCES

- [1] L. Hong and A.K. Jain, "Multimodal Biometrics", in A.K. Jain, R. Bolle and S. Pankanti (Eds.): *Biometrics: Personal Identification in a Networked Society*, Kluwer Academic Publishers, 1999, pp. 326-344.
- [2] A. Ross and A. Jain, "Information Fusion in Biometrics", *Pattern Recognition Letters*, Vol. 24, No. 13, pp. 2115-2125, September 2003.
- [3] J. Kittler and F.M. Alkoot, "Sum Versus Vote Fusion in Multiple Classifier Systems", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 25, No. 1, pp. 110-115, January 2003.
- [4] M. Indovina, U. Uludag, R. Snelick, A. Mink and A. Jain, "Multimodal Biometric Authentication Methods: A COTS Approach", *Proc. MMUA 2003, Workshop on Multimodal User Authentication*, Santa Barbara, CA, December 11-12, 2003, pp. 99-106
- [5] A. K. Jain, K. Nandakumar and A. Ross, "Score Normalization in Multimodal Biometric Systems", *Pattern Recognition*, Vol. 38, No. 12, pp. 2270-2285, December 2005.
- [6] R. Snelick, U. Uludag, A. Mink, M. Indovina, and A. Jain, "Large Scale Evaluation of Multimodal Biometric Authentication Using State-of-the-Art Systems", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 27, No. 3, pp. 450-455, March 2005.
- [7] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, Wiley, New York, 2001.
- [8] S. Ribaric, I.Fratic, "A Matching-Score Normalization Technique for Multimodal Biometric Systems", *Proc. 3rd COST 275 Workshop: Biometrics on the Internet*, Hatfield, UK, 27-28 October 2005, pp. 55-58.
- [9] R. Rodman, D. McAllister, D. Bitzer, L. Cepeda and P. Abbitt, "Forensic Speaker Identification Based on Spectral Moments", *Forensic Linguistics*, Vol. 9, No. 1, pp. 22-43, July 2002.
- [10] S. Ribaric and I.Fratic, "A Biometric Verification System Based on the Fusion of Palmprint and Face Features", *Proc. 4th International Symposium on Image and Signal Processing and Analysis*, Zagreb, Croatia, September 15-17, 2005, pp. 12-17
- [11] S. Ribaric and I.Fratic, "An Online Biometric Authentication System Based on Eigenfingers and Finger-Geometry", *Proc. 13th European Signal Processing Conference*, Antalya, Turkey, September 4-8, 2005.
- [12] S. Ribaric and I. Fratic, "Identification System Based on Eigenpalm and Eigenfinger Features", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 11, pp. 1698-1709, January 2005.
- [13] K. Messer, J. Matas, J. Kittler, J. Luetin and G. Maitre, "XM2VTSDB: The Extended M2VTS Database", *Second International Conference on Audio and Video-based Biometric Person Authentication (AVBPA'99)*, Washington D.C., 1999, pp. 72-77

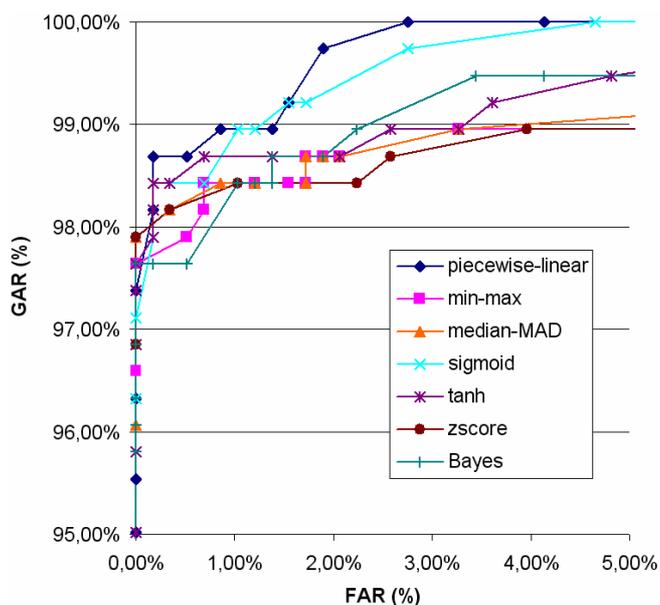


Figure 3. ROC curves for the identification system based on eigenfingers and finger-geometry (I-EF-FG)

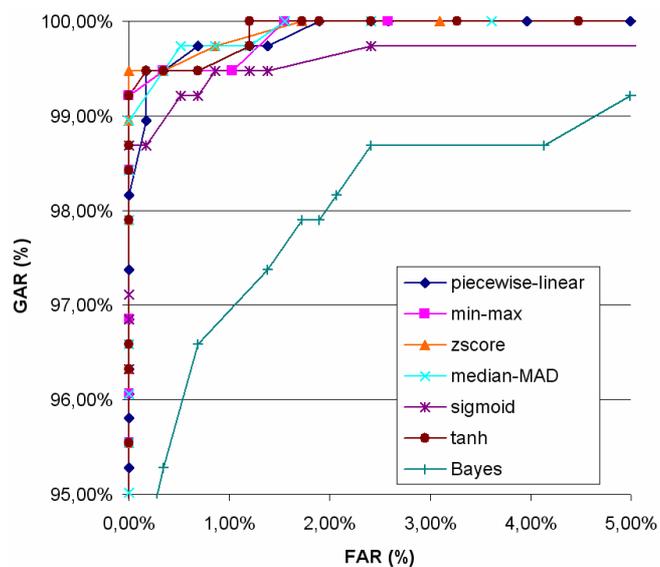


Figure 4. ROC curves for the identification system based on eigenpalm and eigenfinger features (I-EP-EF)