Score Fusion Strategies in Single-Iris Dual-Probe Recognition Systems

P. Drozdowski*†, N. Wiegand*, C. Rathgeb* and C. Busch*

* da/sec - Biometrics and Internet Security Research Group, Hochschule Darmstadt, Germany

† Norwegian Biometrics Laboratory, NTNU, Gjøvik, Norway

{pawel.drozdowski,nikolai.wiegand,christian.rathgeb,christoph.busch}@h-da.de

ABSTRACT

Multiple samples can be utilised at the comparison stage of a biometric system in order to increase its biometric performance via information fusion or decision heuristics. It has been shown, that in a single-instance dual-probe setup, fusing the probe scores yields significant biometric performance increase over the single-probe baseline. Additionally, using the probe-probe comparison score was demonstrated to further improve the biometric performance of a fingerprint recognition system in a study by Cheng et al. In this paper, through a benchmark on the CASIA-IrisV4-Interval dataset and on the iris corpus of the BioSecure dataset, the aforementioned method is shown to be viable for an iris recognition system. However, since it requires an additional parameter, which must be estimated empirically, we propose a simpler method which exhibits similar biometric performance, while requiring no additional parametrisation.

CCS Concepts

• Security and privacy → Biometrics; **• Computing methodologies → Biometrics**;

Keywords

Biometrics, Biometric Information Fusion, Iris Recognition

1 INTRODUCTION

In past years, several multi-biometric iris recognition systems have been proposed [1, 2], some of which consolidate information from multiple samples of a single eye instance during enrolment. Some of these single-instance multi-sample fusion approaches have been found to significantly improve the recognition accuracy of iris recognition systems. The vast majority of proposed iris-based multi-sample fusion schemes process multiple extracted feature vectors, i.e. binary iris-codes, at the time of enrolment. The first conceptual scheme of this kind was presented in [3], in which a majority vote-based coding is applied for each bit position of an odd number of iris-codes, with the goal of reducing the intra-class variation between the resulting reference and probe iris-codes. In [4], a weighted majority voting was proposed to improve the accuracy of an iris recognition system. A weight map, which indicates the stability of iris-code bits, is obtained from several iris-codes at enrolment. Comparison scores are then estimated as a weighted sum of mis-matching bits. A similar approach based on personalized weight maps has been presented in [5]. In [6], so-called “fragile” bits, i.e. bits which exhibit a higher probability than others to flip their value during a genuine comparison, are detected by comparing several iris-codes obtained from a single eye. Incorporating those bits into noise masks extracted in the iris segmentation stage was shown to improve the overall biometric performance of the iris recognition system. In contrast to the aforementioned approaches, a signal-level fusion of iris texture information extracted from multiple frames of a video was proposed in [7]. Based on a pixel-wise averaging, a single normalised iris texture is obtained. Such textures exhibit higher quality/reliabiltiy, and have been shown to improve the biometric performance of an iris recognition system. This scheme has been derived from a concept which was first introduced for face recognition [8]. Similar schemes have been proposed for fingerprint recognition systems [9, 10]. In [11], a score fusion of single-fingerprint dual-probe is proposed, where in addition to utilising the reference-probe comparison scores, the probe-probe comparison score is incorporated into a score fusion. In this paper, said score fusion method, along with proposal of further heuristics are applied in an iris-based system and benchmarked.

The remainder of this paper is organised as follows: in section 2, the fusion strategies for single-iris dual-probe iris recognition are described. In section 3, the experimental setup and results are presented, while section 4 contains a summary of the paper.

2 FUSION STRATEGIES

State-of-the-art iris recognition systems capture multiple samples during acquisition stage for the purpose of supporting compensation of pose or gaze variations or for providing some fundamental presentation attack detection (PAD) [12]. Those additional samples can then be utilised at comparison stage. Specifically, in a system where two probe samples are present at comparison stage, three comparison scores can be computed as shown in figure 1: two (HD_1 and HD_2) between the reference and each probe and one (HD_3) between the two probes themselves. It is then possible to fuse the scores, for example, in following ways:

- Using only the scores between the reference and probes, an Average Reference-Probe score (referred to as "ARP"): $(HD_1 + HD_2)/2$. Observe, that for fusing the scores no normalisation is required, since the experimental scores stem from a single biometric system (same modality and same comparison algorithm).
- Using all three scores, an Average Reference-Probe score weighted by the probe-probe score (referred to as "w-ARP"): $(HD_1 - a * HD_3 + HD_2 - a * HD_3)/(2 - a)$, where a is estimated on a training set, so that it maximises the biometric performance.
- Based on the probe-probe score, the reference-probe scores are either fused using ARP or only the minimum is used (referred to as "Min-or-ARP"). Here, the probe-probe score functions as a quality check – if one (or both) probes are of bad quality, then HD_3 is likely to be high. In this case, if HD_3 exceeds the acceptance threshold of the biometric system, it will therefore be better, instead of ARP, to simply use the

Post-print: ICBEA ’18, May 16–18, 2018, Amsterdam, Netherlands
https://doi.org/10.1145/3230820.3230823

minimum of HD_1 and HD_2. Doing so will disproportionally favour genuine transactions, by providing better chances of acceptance even in case of one sample being of bad quality; whereas the impact on impostor scores is expected to be negligible.

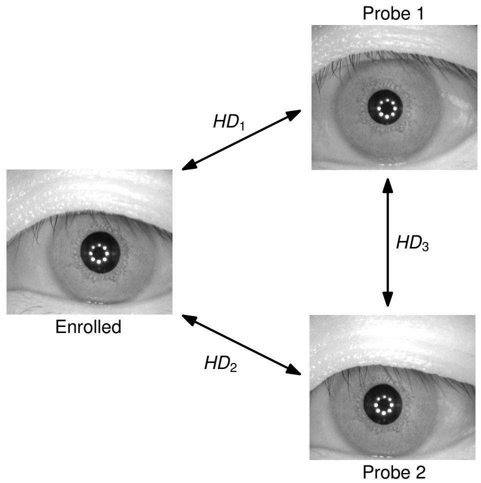


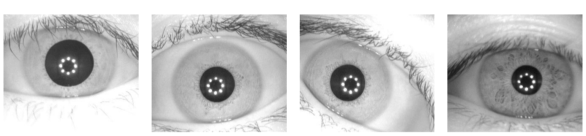
Figure 1: Single-iris dual-sample iris recognition

3 PERFORMANCE EVALUATION

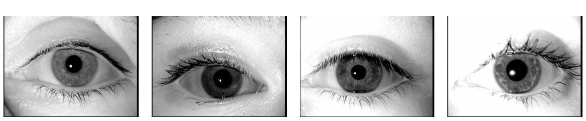
This section contains the evaluation of the dual-sample fusion schemes described in section 2. In subsection 3.1, the used dataset and the experimental setup details are outlined, while the results are presented and discussed in subsection 3.2.

3.1 Dataset

The experiments were performed on the CASIA-IrisV4-Interval database [13] (henceforth referred to as "CASIA") and the iris corpus of the BioSecure database [14] (henceforth referred to as "BioSecure"), both containing images captured in near-infrared light spectrum. An overview of the datasets is shown in table 1, while example images are shown in figure 2. Several subjects had to be removed from the BioSecure dataset due to labelling errors.



(a) CASIA



(b) BioSecure

Figure 2: Example images from the datasets

The raw images were processed with the commonly used methods, as shown in figure 3. After segmentation using the Viterbi algorithm [15], where the iris and pupil boundaries are located, the iris textures were normalised according to the rubbersheet model [16] and subsequently enhanced by applying Contrast Limited

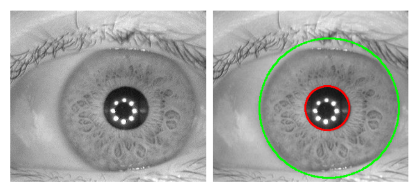
Table 1: Dataset overview

Dataset	Subjects	Instances	Images	Resolution
CASIA	249	395	2639	320 × 280 px
BioSecure	210	420	1680	640 × 480 px

Table 2: Numbers of comparisons performed during experiments **("Fusion" refers to all three fusion experiments, i.e. ARP, w-ARP and Min-or-ARP, since for each one of those the transactions numbers are identical)**

Dataset	Experiment	Genuine	Impostor
CASIA	Baseline	41594	7993888
	Fusion	20797	3996944
BioSecure	Baseline	10080	2111632
	Fusion	5040	1055816

Adaptive Histogram Equalization (CLAHE). Feature extraction was performed with the Daugman-like 1D-LogGabor algorithm (LG), generating iris-codes of size 512 × 20 = 10240 bits. Such templates are compared using fractional Hamming distance with circular shifts applied to account for sample misalignment. The implementations of the aforementioned algorithms were provided by open-source frameworks OSIRIS [17] and USIT [18]. The evaluation of the methods described in section 2 along with a single-sample baseline were performed in verification mode. In the experiments, all possible transactions were performed; table 2 shows the numbers of transactions for each experiment.



(a) Iris detection in a raw image

(b) Normalised and enhanced texture

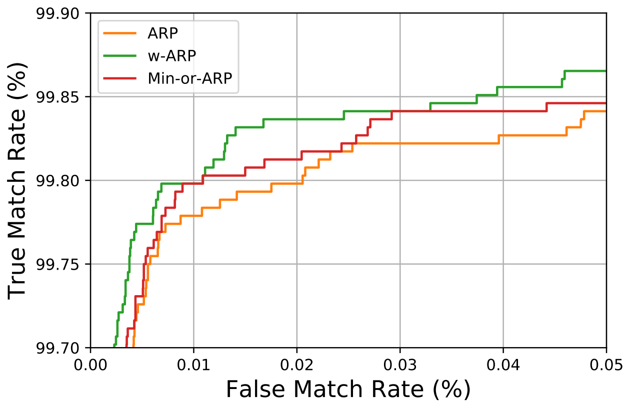
(c) 1D Log-Gabor Iris-Code

Figure 3: Iris recognition processing chain

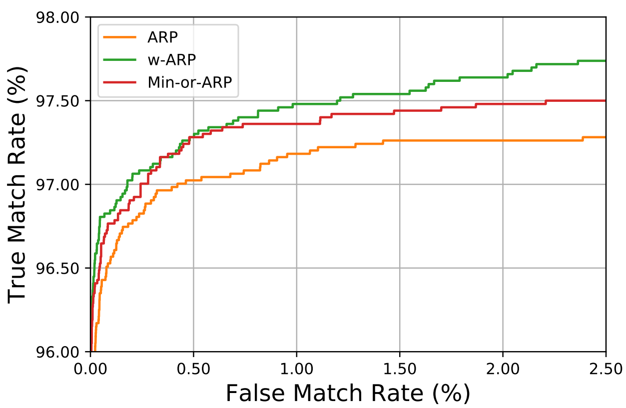
3.2 Results

Figure 4 shows the receiver operating characteristic (ROC) curves of the benchmarked approaches.

The baseline is not shown, since its biometric performance is well below that of the fusion approaches (see table 3). It can be seen, that by incorporating the third comparison score (between the two probes – w-ARP) into the score fusion, biometric performance can be improved over that of a simple score fusion of the two reference-probe scores (ARP). It also appears that said third comparison score can be effectively used as a quality check, since the Min-or-ARP algorithm slightly outperforms the plain score fusion strategy (ARP) and has a biometric performance comparable to that of w-ARP. In figure 5, it can be seen that the biometric performance of w-ARP

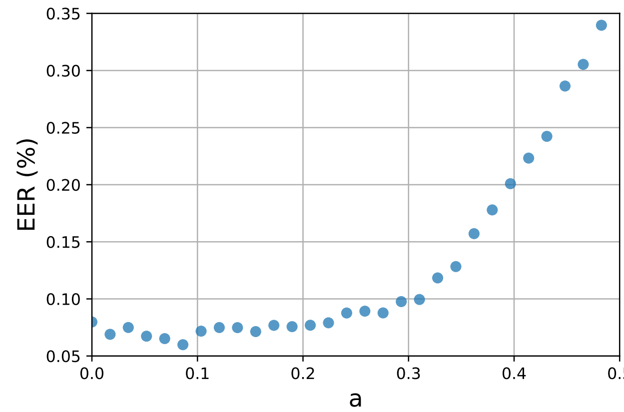


(a) CASIA

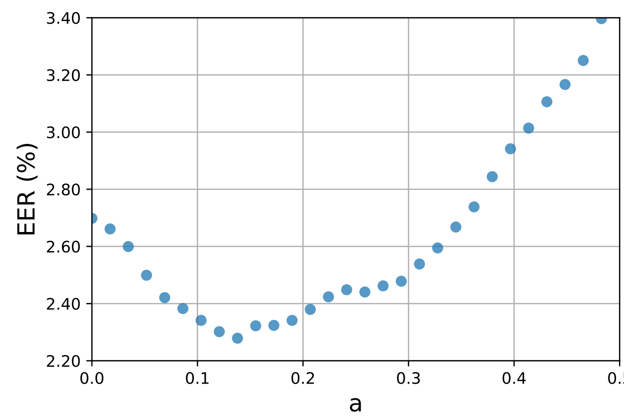


(b) BioSecure

Figure 4: ROC curves



(a) CASIA



(b) BioSecure

Figure 5: Scatter plots for w-ARP scheme showing the dependence of biometric performance on the a parameter

varies strongly with the value of the a parameter. This provides strong motivation for introducing the Min-or-ARP scheme, as a parameter-free alternative.

In table 3, additional metrics for benchmarking the strategies are listed. Those are: equal-error-rate (EER), area under ROC curve (AUC) and decidability (d'). The decidability is computed using the means and standard deviation of the genuine and impostor score distributions: $d' = \frac{|\mu_1 - \mu_2|}{\sqrt{\frac{1}{2} * (\sigma_1^2 + \sigma_2^2)}}$ (higher values are better). This metric is useful in assessing the intrinsic decidability of a biometric decision problem, although with the limitation of ignoring statistical moments higher than second-order [19]. It can be observed, that the dual-sample set-ups all outperform the baseline significantly, while the benefits of the additional heuristics (w-ARP, Min-or-ARP) over ARP are noticeable, especially in the significantly higher decidability values.

In addition to the biometric performance and decidability metrics, it is interesting to take a look at the statistical and visual properties of the genuine and impostor score distributions produced by the algorithms described in section 2. Those are listed in table 4, while figure 6 shows kernel density estimates of the distributions. Most noticeable is that the genuine distribution for w-ARP has significantly shifted to the left, while its corresponding distribution

Table 3: Results

Dataset	System	EER	AUC	d'
CASIA	Baseline	0.00334	0.99938	5.88276
	ARP	0.00130	0.99968	6.69575
	w-ARP	0.00110	0.99969	7.07860
	Min-or-ARP	0.00115	0.99968	6.74932
BioSecure	Baseline	0.04563	0.97855	3.80808
	ARP	0.02698	0.98806	4.25278
	w-ARP	0.02317	0.99103	4.60225
	Min-or-ARP	0.02455	0.98919	4.49606

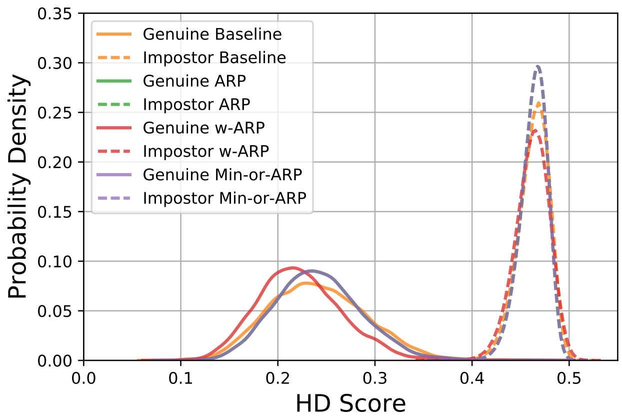
has only done so slightly, which explains the improved biometric performance.

4 SUMMARY

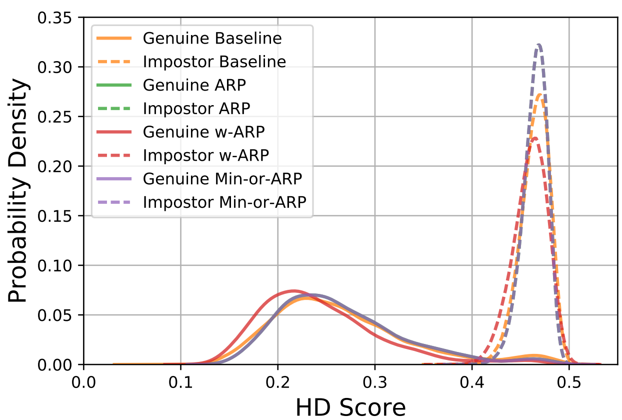
In this paper, several methods for fusing information in single-iris dual-probe authentication scenario were benchmarked. It has been shown that using two probe samples can yield significant biometric performance improvements over the single probe sample baseline. Specifically, aside from a simple score fusion of the two reference-probe scores, a third score – between the two probes – can be utilised. Here, two methods were tested: one a direct

Table 4: Distribution statistics

Dataset	Type	System	Min	Max	Mean	Std	Skew	Ex. kurt.
CASIA	Genuine	Baseline	0.076	0.484	0.242	0.051	0.377	0.390
		ARP	0.090	0.484	0.242	0.045	0.383	0.506
		w-ARP	0.079	0.487	0.223	0.044	0.532	0.821
		Min-or-ARP	0.090	0.484	0.242	0.044	0.363	0.496
	Impostor	Baseline	0.354	0.524	0.463	0.016	-0.568	0.479
		ARP	0.371	0.520	0.463	0.014	-0.599	0.587
		w-ARP	0.355	0.530	0.461	0.018	-0.397	0.277
		Min-or-ARP	0.355	0.520	0.463	0.014	-0.607	0.616
BioSecure	Genuine	Baseline	0.064	0.497	0.266	0.072	0.972	0.846
		ARP	0.142	0.491	0.266	0.065	0.959	0.875
		w-ARP	0.120	0.497	0.248	0.063	1.078	1.415
		Min-or-ARP	0.142	0.491	0.262	0.062	1.031	1.304
	Impostor	Baseline	0.351	0.526	0.465	0.015	-0.591	0.499
		ARP	0.370	0.511	0.465	0.013	-0.631	0.669
		w-ARP	0.356	0.522	0.460	0.017	-0.395	0.087
		Min-or-ARP	0.370	0.511	0.464	0.013	-0.639	0.667



(a) CASIA



(b) BioSecure

Figure 6: Kernel density estimates for the score distributions

re-implementation of the idea from single-fingerprint dual-probe system of Cheng et al., where the third score is directly incorporated into the score fusion. A second method was proposed, where the third score acts as a probe quality check, based on which the reference-probe scores are either fused or only the minimum is used. Both methods yield slight improvements over the simple score fusion method in terms of biometric performance (ROC curves) and decidability (d'). The advantage of the proposed method (Min-or-ARP) over the existing weighted fusion method (w-ARP) is that it does not require additional parametrisation (in w-ARP, the a parameter has to be estimated on a training set to minimise the EER).

For the systems operating in verification mode, the additional computational workload of the dual-sample approach is negligible – 2 or 3 template comparisons instead of 1, while in the identification mode, the workload would be doubled (the score between the two probes need only be calculated once). Lastly, the dual-sample approach could be effortlessly incorporated into some operational systems, since they might already capture multiple samples, e.g. for PAD.

ACKNOWLEDGEMENTS

This work was partially supported by the German Federal Ministry of Education and Research (BMBF) as well as by the Hessen State Ministry for Higher Education, Research and the Arts (HMWK) within Center for Research in Security and Privacy (CRISP).

REFERENCES

[1] P. Radu, K. Sirlantzis, G. Howells, F. Deravi, and S. Hoque. 2012. A Review of Information Fusion Techniques Employed in Iris Recognition Systems. *Int. J. Adv. Intell. Paradigms* 4, 3/4 (February 2012), 211–240.

[2] A. Ross, K. Nandakumar, and A. K. Jain. 2006. *Handbook of multibiometrics.* Springer Science.

[3] G. Davida, Y. Frankel, and B. Matt. 1998. On enabling secure applications through off-line biometric identification. In *Proc. IEEE Symp. on Security and Privacy.* IEEE, 148–157.

[4] S. Ziauddin and M. N. Dailey. 2008. Iris recognition performance enhancement using weighted majority voting. In *Intl. Conf. on Image Processing (ICIP).* 277–280.

[5] W. Dong, Z. Sun, and T. Tan. 2011. Iris Matching Based on Personalized Weight Map. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 33, 9 (September 2011), 1744–1757.

[6] K. P. Hollingsworth, K. W. Bowyer, and P. J. Flynn. 2009. The Best Bits in an Iris Code. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 31, 6 (June 2009), 964–973.

[7] K. Hollingsworth, T. Peters, K. W. Bowyer, and P. J. Flynn. 2009. Iris Recognition Using Signal-Level Fusion of Frames From Video. *IEEE Transactions on Information Forensics and Security* 4, 4 (December 2009), 837–848.

[8] R. Chellappa, V. Kruger, and Shaohua Zhou. 2002. Probabilistic recognition of human faces from video. In *Intl. Conf. on Image Processing (ICIP)*, Vol. 1. I–41–I–44 vol.1.

[9] A. Jain and A. Ross. 2002. Fingerprint mosaicking. In *Intl. Conf. on Acoustics, Speech, and Signal Processing*, Vol. 4. IV–4064–IV–4067.

[10] C. Ryu, Y. Han, and H. Kim. 2005. Super-template Generation Using Successive Bayesian Estimation for Fingerprint Enrollment. In *Intl. Conf. on Audio- and Video-Based Biometric Person Authentication (AVBPA'05).* Springer, 710–719.

[11] X. Cheng, S. Tulyakov, and V. Govindaraju. 2011. Multiple-sample fusion of matching scores in biometric systems. In *Computer Vision and Pattern Recognition Workshops (CVPRW).* 120–125.

[12] J. Galbally and M. Gomez-Barrero. 2016. A review of iris anti-spoofing. In *Intl. Conf. on Biometrics and Forensics (IWBF).* 1–6.

[13] Chinese Academy of Sciences' Institute of Automation. 2010. CASIA iris image database. http://biometrics.idealtest.org/. (December 2010). Last accessed: 2017-10-11.

[14] J. Ortega-Garcia, J. Fierrez, F. Alonso-Fernandez, J. Galbally, M. R. Freire, et al. 2010. The Multiscenario Multienvironment BioSecure Multimodal Database (BMDB). *IEEE Trans. on Pattern Analysis and Machine Intelligence* 32, 6 (June 2010), 1097–1111.

[15] G. Sutra, S. Garcia-Salicetti, and B. Dorizzi. 2012. The Viterbi algorithm at different resolutions for enhanced iris segmentation. In *Intl. Conf. on Biometrics (ICB).* IEEE, 310–316.

[16] J. Daugman. 2004. How iris recognition works. *IEEE Trans. on Circuits and Systems for Video Technology* 14, 1 (January 2004), 21–30.

[17] N. Othman, B. Dorizzi, and S. Garcia-Salicetti. 2016. OSIRIS: An open source iris recognition software. *Pattern Recognition Letters* 82, 2 (September 2016), 124–131.

[18] C. Rathgeb, A. Uhl, P. Wild, and H. Hofbauer. 2016. Design Decisions for an Iris Recognition SDK. In *Handbook of Iris Recognition* (2nd ed.), K. Bowyer and M. J. Burge (Eds.). Springer.

[19] J. Daugman. 2000. *Biometric Decision Landscapes.* Technical Report UCAM-CL-TR-482. University of Cambridge - Computer Laboratory. 15 pages.