Segmentation-level Fusion for Iris Recognition

Peter Wild[*‡], Heinz Hofbauer[†], James Ferryman[*] and Andreas Uhl[†]

[*]School of Systems Engineering, University of Reading, Reading RG6 6AY, UK.

`j.m.ferryman@reading.ac.uk`

[†]Dept. of Computer Sciences, University of Salzburg, 5020 Salzburg, Austria.

`{hhofbaue, uhl}@cosy.sbg.ac.at`

[‡]AIT Austrian Institute of Technology GmbH, 2444 Seibersdorf, Austria.

`peter.wild@ait.ac.at`

Abstract—This paper investigates the potential of fusion at normalisation/segmentation level prior to feature extraction. While there are several biometric fusion methods at data/feature level, score level and rank/decision level combining raw biometric signals, scores, or ranks/decisions, this type of fusion is still in its infancy. However, the increasing demand to allow for more relaxed and less invasive recording conditions, especially for on-the-move iris recognition, suggests to further investigate fusion at this very low level. This paper focuses on the approach of multi-segmentation fusion for iris biometric systems investigating the benefit of combining the segmentation result of multiple normalisation algorithms, using four methods from two different public iris toolkits (USIT, OSIRIS) on the public CASIA and IITD iris datasets. Evaluations based on recognition accuracy and ground truth segmentation data indicate high sensitivity with regards to the type of errors made by segmentation algorithms.

Fig. 1: Iris Segmentation Fusion Framework.

I. INTRODUCTION

Iris recognition challenges for on-the-move and less constrained acquisitions, like the Noisy Iris Challenge Evaluation (NICE) [1], and Multiple Biometrics Grand Challenge (MBGC), illustrated the importance of robust iris segmentation in latest-generation iris biometric systems. Iris verification rates as low as 44.6% [2] are reported for unconstrained applications, and image quality has been shown to play a critical role in the segmentation and normalisation process [3]. Normalisation seems to be at the heart of the problem, but combination past feature-extraction (see fusion scenarios in [4]) is easier and segmentation fusion lacks standardisation. ISO/IEC TR 24722:2007 does not foresee multinormalisation, nor does ISO/IEC 19794-6:2011 define a segmentation-only exchange format: there is a *cropped and masked* data format for normalised textures following IREX K7, but without direct access to segmentation results/parameters. While segmentation algorithms themselves might combine different approaches, iris segmentation fusion as proposed in [5] is widely ignored as a means to achieve more robust and accurate segmentation. As a common alternative, multi-algorithm fusion is suggested as a scenario [4] operating on the same input images. However, the expected increase in accuracy is usually not justifying the cost (in terms of additional processing power). Strong correlation of algorithms combined at system levels due to similar/same normalisation steps, and the dominance of local Gabor-based features (following Daugman's rubbersheet normalisation and original feature extraction [6]) are likely to be reasons for observed little impact on accuracy (compared to combining, e.g. image-based and binary features [7]). Fusion at image data level, such as in [8] following [9] reveals promising results, but requires the multiple execution of the iris unwrapping
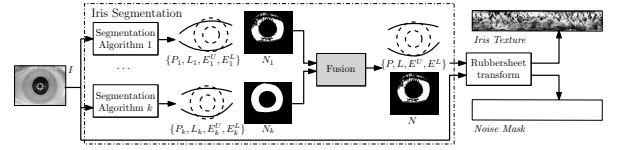
and normalisation process (for each obtained segmentation). Furthermore, given multiple normalised textures after segmentation and unwwrapping, it is difficult to determine faulty or highly inaccurate segmentation versions.

The novelty of this work is a thorough analysis of how segmentation-based fusion in iris recognition can help in achieving higher accuracy considering the entire iris processing chain involving feature extraction, which may itself be tolerant to deformation to a certain extent. The latter observation raises the question on evaluation of fusion schemes at this stage, as ground-truth conformance is just one of several impacting factors. Especially the impact of outliers is highlighted in this paper. For this task, (1) a framework of combining segmentation results following Daugman's rubbersheet model is presented (see Fig. 1); (2) a set of reference fusion methods combining segmentation curves, models, and masks is implemented, and; (3) pairwise combination improvement is analysed on public datasets with regards to both, ground-truth and recognition-acuracy. The following questions are addressed in this paper: (1) Does the combination of automated iris segmentation results yield more accurate result than each of the employed original segmentation algorithms? (2) How does the choice of database and segmentation algorithms impact on iris segmentation fusion? (3) How do outliers impact on overall recognition accuracy and how do ground-truth-based vs. recognition-based evaluations relate to each other?

As an introduction to the topic of multi-segmentation fusion Section II reviews related work on iris normalisation, fusion approaches, and segmentation data interoperability. Section III presents the proposed framework of segmentation fusion and discusses implementations. An experimental evaluation of proposed techniques is given in Section IV, analysing results with regards to questions outlined in this introduction. Finally, Section V concludes this work on segmentation-based fusion for iris biometric systems.

II. RELATED WORK

Modern iris recognition algorithms operate on normalised representations of the iris texture obtained by mapping the area

between inner and outer iris boundaries $P, L : [0, 2\pi] \to [0, m] \times [0, n]$ to "Faberge" or "Rubbersheet" coordinates" (using angle θ and pupil-to-limbic radial distance r) [6], independent of pupillary dilation: $R(\theta, r) := (1-r) \cdot P(\theta) + r \cdot L(\theta)$. Normalised texture and noise masks $T, M : [0, 2\pi] \times [0, 1] \to C$ are obtained (C is the target color space, $M = N \circ R, T = I \circ R$ for the original $n \times m$ image I and noise mask N). The latter usually considers reflections and upper and lower eyelid curves masking out occlusions, such that $N(x, y) \neq 0$ if and only if pixel (x, y) refers to an in-iris location. While normalisation is standardised, there are several iris segmentation approaches for obtaining P, L and N. Original approaches employed circular boundary-based segmentation, such as Daugman's integro-differential operator [6] and Wildes' circular Hough Transform (HT) [10]. Today's advanced iris segmentation techniques are often multi-stage approaches combining various techniques: Active shape models [11], clustering-based iris localization [12] (e.g. locating the sclera for NICE.I data), AdaBoost-cascade and Pulling-and pushing models [13], agent-based methods [14], the Viterbi algorithm at different resolutions [15], or iterative multi-scale approaches and ellipsopolar transform for elliptical iris models [16]. With the recent focus on visible-range (VR) iris segmentation compared to traditional near-infrared (NIR) segmentation techniques, the robust combination of independent segmentation approaches becomes an interesting aspect. Recently, [17] compared multiple segmentation algorithms on different VR and NIR datasets based on ground truth information, illustrating the dependence of algorithms on database-specific assumptions and underlining the need for more robust segmentation.

There are not many proposed fusion techniques operating before feature extraction, most of them focusing on data-level fusion: Huang et al. [18] present a Markov network learning-based fusion method to enhance the resolution of iris images. Hollingsworth et al. [19] combine high-resolution images from multiple frames to create a single combined representation. Jillela and Ross [20] proposed image-level fusion with Principal Components Transform. Recently, Llano et al. [8] investigate the positive segmentation impact of PCA-based fusion vs. Laplacian Pyramid and Exponential Mean at image-level, i.e. multiple normalised iris textures are fused retrieved by following different segmentation algorithms. A first fusion approach of segmentation information (i.e. prior to normalisation) with the benefit of single normalisation and potentially simpler treatment and classification of errors than post-normalisation fusion is proposed in [5].

This work builds upon the framework of fusion for multiple iris segmentations introduced by Uhl and Wild [5], who combined evidence from human (manual) ground truth segmentation as a proof of concept work, but without any tests on automated iris segmentation algorithms and on a single dataset only. Two fusion methods were tested, both achieved higher recognition accuracy independent of the employed feature extraction algorithm (testing 3 approaches). Yet, the type of fusion technique (model-wise or data-wise) did not have a huge impact on accuracy and manual segmentation was reported to be fairly stable with 97.46% vs. 97.64% genuine acceptance rate (GAR), at 0.01% false acceptance rate (FAR), without any severe segmentation outliers [5].

The performance on automated segmentation algorithms

raises further questions, especially questions related to stability if algorithms fail: Accurate results of a cohort of segmenters might be severely affected by a single segmentation error. Further, evaluations will be extended to ground-truth segmentation information as suggested by the Noisy Iris Challenge Evaluation - Part I (NICE.I), and the F-measure used in [17]: Errors are estimated from the segmentation result (noise mask) N_i (or, more specifically, to estimate boundary detection performance an artificial noise mask is constructed rendering the iris using boundary curves P_i, L_i) for each image I_i and compared using a ground truth mask G_i. Let tp_i, fp_i, tn_i, fn_i refer to true / false respectively positive / negative pixel in-iris classifications for image index i (with dimension $m \times n$), then:

$$E_1 := \frac{1}{k} \sum_{i=1}^{k} \frac{fp_i + fn_i}{mn}; \qquad (1)$$

$$E_2 := \frac{1}{2} \left(\frac{1}{k} \sum_{i=1}^{k} \frac{fp_i}{fp_i + tn_i} \right) + \frac{1}{2} \left(\frac{1}{k} \sum_{i=1}^{k} \frac{fn_i}{fn_i + tp_i} \right) \qquad (2)$$

$$\text{F-measure} = F_1 := \frac{1}{k} \sum_{i=1}^{k} \frac{tp_i}{tp_i + \frac{1}{2}(fn_i + fp_i)} \qquad (3)$$
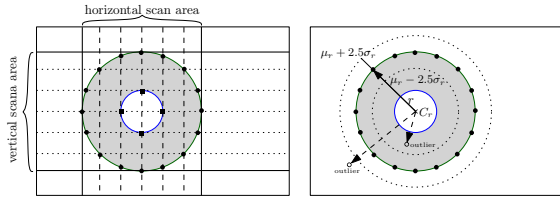
Error rate E_1 refers of the rate of pixel disagreement between ground truth and segmentation noise masks, E_2 accounts for the disproportion between a priori probabilities, F_1 gives a measure of correctly to incorrectly proportions. Augmenting [5], this paper evaluates ground truth accuracy (using public IRISSEG-EP [17]) and recognition impact, including an exhaustive significance analysis, to gain a deeper understanding of reasons for improvement. The McNemar test [21] is used for statistical significance analysis. In contrast to [8] this work does not assume access to multiple source images and unlike [9] does not rely on multiple normalisations. However, the same open segmentation and recognition algorithms (USIT) are employed for reproducibility.

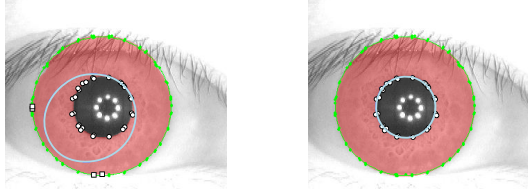### III. MULTI-SEGMENTATION FUSION METHODS

Modern iris recognition algorithms pushed by challenge measures (NICE.I E_1, E_2 as introduced in Sect. II) focus on the problem of boundary refinement, taking occlusions and reflections into account [11], [12], [15]). For the Faberge mapping however, a robust segmentation of true (potentially occluded) boundaries P, L is critical, neglecting the presence of noise artifacts. This is to avoid non-linear distortions [16]. While such distortions could possibly be targeted by more sophisticated matching techniques (e.g. by using Levenshtein distance), in identification mode it is more time-efficient to employ fast matching and study more advanced normalisation techniques, or combinations thereof and subject to this paper. In case direct parameterisations of the algorithm are available (e.g. center and radius for circular models, elliptical models, splines, or polygonal boundary representations), the following techniques have been proposed in [5]:

$$\text{Sum Rule: } B(\theta) := \frac{1}{k} \sum_{i=1}^{k} B_i(\theta); \qquad (4)$$

$$\text{Aug Rule: } B(\theta) := ModelFit \left(\bigcup_{i=1}^{k} B_i \right)(\theta) \qquad (5)$$

(a) Overview over the iris scanning and pruning process.

(b) With outliers   (c) With outliers pruned

Fig. 2: Overview over the iris scanning and pruning process with examples.

- **Sum-Rule Interpolation**: This fusion rule combines boundary points $B_i(\theta)$ of curves $B_1, B_2, \ldots B_k$: $[0, 2\pi) \to [0, m] \times [0, n]$ into a single boundary B, for pupillary and limbic boundaries, in analogy to the sum rule.

- **Augmented-Model Interpolation**: This model combines boundaries $B_1, \ldots, B_k$ within a jointly applied parametrisation model *ModelFit* minimizing the model-error (e.g., Fitzgibbon's ellipse- [22], or least-squares circular fitting), executed separately for inner and outer iris boundaries. Models are combined, not only points.

Segmentation masks N are common intermediate results, as normalisation is usually integrated rather than providing parameterisations of boundaries. A natural approach therefore is the extraction of parameterisations using noise masks, e.g. by employing an elliptical fitting. The following section illustrates the mask scanning process proposed in this work.

A. Scanning Iris Masks

The mask fusion is an augmented-model interpolation based on a scan of the provided iris masks. This fusion method is based on the assumption that the mask is available but the original fitted model for the iris, pupil and eyelid boundaries are not, as would be the case for the IREX K7 specification. It follows the basic outline of the augmented model interpolation but skips the eyelid polygon fits. In a scan it is not necessarily possible to differentiate between iris and eyelid based purely on the mask. The model used for the augmentation is an ellipse fitting based on a scan of the iris mask.

First, the iris boundaries for each axis are determined. Then N equidistant scan lines are used to generate points along the iris and pupil boundaries. The boundary points of the provided masks are combined and pruned for outliers. Outliers typically happen when the outer mask of an iris is not convex, leading to wrongfully detected pupil boundary points along the iris boundary. The outlier detection is done by using the center of gravity C_r of all the detected points for a given boundary. The radius for each point from C_r is calculated and all points are pruned for which the radius has a z-score of greater than

(a) original   (b) correct boundaries   (c) without noise

Fig. 3: Original OSIRIS mask, corrected version for rubber-sheet mapping, and corrected without noise masking.

2.5. This is illustrated in Fig. 2 along with the difference of pruned and unpruned iris detection. In order to get a stable outlier detection and correct boundaries a high number of scan lines is desirable, for our experiments $N = 100$ was used. Furthermore, to properly associate mask transitions with iris or pupil boundaries there should be no extra transitions. Such transitions can be generated by noise exclusions in the mask. If the mask contains holes of this kind, they should be closed by an dilate+erode morphological operation.

The implementation of the tested OSIRIS algorithm produces masks which extend over the actual boundaries used for unrolling the iris image, see Fig. 3a, which would produce incorrect masks during the scanning steps. For the experiments we modified OSIRIS to restricted the produced mask to the detected boundaries, see Fig. 3b. In addition, we introduced an option to skip the noise mask, resulting in masks as shown in Fig. 3c. This speeds up the fusion of the masks by allowing to skip the dilate+erode morphological operations. The noise mask is only skipped for mask fusion not for calculation of the OSIRIS scores.

The mask fusion algorithm produces two points clouds, pertaining to the iris and pupil boundary. The actual mask is generated by fitting an ellipse to the point clouds by a least-squares method [22]. The segmentation tool from [17] is used for unrolling the iris image. It should also be noted that the mask level fusion only generates a mask which is used for unrolling the iris. No noise or occlusion mask is generated and consequently all tests performed on the fusion are performed purely on the unrolled iris image without masking.

IV. EXPERIMENTAL STUDY

Addressing the question of multisegmentation fusion performance, we assessed pairwise combinations of the following segmentation algorithms: CAHT [23], a traditional sequential (limbic-after-pupillary) method based on circular HT and contrast-enhancement; WAHET [16], a two-stage adaptive multi-scale HT segmentation technique using elliptical models; OSIRIS [24], a circular HT-based method with boundary refinement; IFPP [25] using iterative Fourier-series approximation and Pulling and Pushing methods. The motivation for selecting these algorithms were public availability as open source software for reproducibility, therefore also basing experiments on ground-truth segmentations released with [17], [26] and referring to the public CASIA-v4 and IITD iris databases. As feature extractors we used the wavelet transform-based algorithm by Ma *et al.* [27] and the local Gabor-filter based algorithm by Masek [28]. The results in terms of equal error rate were obtained by using Hamming distance based

TABLE I: Equal error rate for the segmentation fusion.

(a) Casia v4 Interval database

Equal-error rate [%] of Masek			
	CAHT	WAHET	OSIRIS	IFPP
CAHT	1.22	**0.92**	**1.03**	1.30
WAHET		1.89	**1.02**	**1.41**
OSIRIS			1.04	1.44
IFPP				8.10

Equal-error rate [%] of Ma			
	CAHT	WAHET	OSIRIS	IFPP
CAHT	0.99	**0.64**	0.84	1.17
WAHET		1.72	0.89	**1.22**
OSIRIS			0.73	1.53
IFPP				8.78

(b) IIT Delhi database

Equal-error rate [%] of Masek			
	CAHT	WAHET	OSIRIS	IFPP
CAHT	1.85	3.60	1.65	**1.38**
WAHET		6.82	3.90	**3.70**
OSIRIS			1.40	1.94
IFPP				3.87

Equal-error rate [%] of Ma			
	CAHT	WAHET	OSIRIS	IFPP
CAHT	1.72	4.06	*1.95*	**1.43**
WAHET		7.43	4.86	**4.23**
OSIRIS			1.21	2.40
IFPP				4.36

TABLE II: Results of the McNemar test, reported as the X^2 values. The row gives the single method compared to the fusion as indicated by row×column.

(a) Casia v4 Interval database

X^2 statistic for Masek			
	single method			
fused with	CAHT	WAHET	OSIRIS	IFPP
CAHT		24742	8	246149
WAHET	2543		13	247450
OSIRIS	1158	22002		243734
IFPP	928	8110	3729	

X^2 statistic for Ma			
	single method			
fused with	CAHT	WAHET	OSIRIS	IFPP
CAHT		28739	135	273347
WAHET	3993		1649	276351
OSIRIS	1620	15752		261445
IFPP	1438	7076	10532	

(b) IIT Delhi database

X^2 statistic for Masek			
	single method			
fused with	CAHT	WAHET	OSIRIS	IFPP
CAHT		49180	169	35918
WAHET	20317		42328	24
OSIRIS	1746	27835		17116
IFPP	3193	38721	3655	

X^2 statistic for Ma			
	single method			
fused with	CAHT	WAHET	OSIRIS	IFPP
CAHT		21271	4614	61327
WAHET	52945		78177	53
OSIRIS	368	10149		26311
IFPP	1145	21256	11669	

verification, the tools are provided by and further documented in the USIT package [23].

A. Impact on Recognition Accuracy

The main motivation for combining segmentation algorithms is to achieve a better overall system recognition accuracy. Whereas segmentation is an integral part of a biometric recognition system, the advantage of a system-based evaluation is that it takes into account that small segmentation errors do not necessarily implicate an impact on recognition accuracy, as the feature extraction (and comparison) algorithm itself tries to extract features invariant under slight transformations (e.g. small shifts, different illumination, etc.). Table Ia gives the results of the evaluation on the CASIA-IrisV4-Interval [29] database, and Table Ib gives the results on the IIT Delhi Iris Database [30]. The entries along the principal diagonal are the results of the original segmentation algorithms. Fusion results which are an improvement over both fused algorithms are shown in a bold font and fusion results where the fusion performs worse than both individual algorithms are shown in italics.

From Table I we can see that segmentation fusion increased performance in 10 out of 24 combination scenarios involving different algorithms and databases. While there is only one case, IFPP fused with WAHET, which consistently increases the performance, there are numerous cases where the fusion improves over both algorithms. In particular, there is only one case, OSIRIS fused with CAHT with feature extraction of Ma on the IITD database, where the combined performance is worse than both solitary performances. Given that all employed segmentation algorithms aim for gradient-based detection rather than employing completely different approaches and thus limiting the fusion potential as any independence assumption is likely violated, the fraction of cases with improvement is rather encouraging and deserves further attention.

In order to verify the statistical significance of results, we conducted McNemar tests [21] dedicated to matching pairs of subjects. The test uses the dichotomous trait of correct classfication (in relation to the known ground truth). We utilize the χ^2 approximation with the continuity correction proposed by Edwards [31]. Table II reports obtained X^2 values at the EER operating point. Note that a critical value $X^{2*} \geq 6.64$

indicates a rejection of the null hypothesis — that there is no difference between the two methods — with at least 99% significance. The table gives the comparison of single method in the column, e.g. CAHT(column), with the fusion as indicated by column and row, e.g. CAHT(column) fused with WAHET(row).

B. Ground-truth Segmentation Accuracy

To understand how fusion influences the segmentation performance we compared the segmentation results to ground truth, which is available from two independent manual segmentations. Fig. 4 gives the F-measure segmentation error introduced in Eq. 3 for IFPP, WAHET and their fusion on the CASIA v4 interval database. The fusion exhibits a closer conformity to the ground truth than each individual segmentation algorithm. Using the outlier detection from [17] we can further confirm the conformity to the ground truth for the fusion; IFPP had 95 outliers, WAHET had 32 and the fusion only 16.

We compared Sum Rule segmentation fusion performance on "good" versus "bad" segmentations using segmentation consistence between both algorithms as an indicative measure (we used distance of pupillary and limbic centers, as well as the absolute difference in radii using threshold $\eta = 10$). Results indicated, that fusion performance on the "good" set improved

TABLE III: Fusion for good vs. bad segmentation results.

Segmentation error [%]				
	E_1		E_2	
	Good	Bad	Good	Bad
CAHT	1.98	2.76	3.02	4.10
WAHET (NIR)	2.30	6.05	3.54	8.90
Fusion (Sum Rule)	1.87	3.85	2.87	5.61

(a) Casia v4 Interval database

Segmentation error [%]				
	E_1		E_2	
	Good	Bad	Good	Bad
CAHT	2.61	5.00	3.48	8.33
WAHET (NIR)	2.77	15.31	3.73	20.76
Fusion (Sum Rule)	2.40	9.95	3.23	13.84

(b) IIT Delhi database

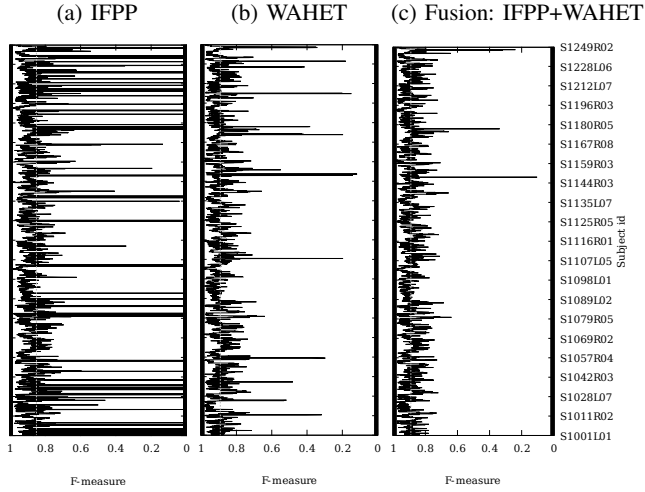(a) IFPP (b) WAHET (c) Fusion: IFPP+WAHET

Fig. 4: Segmentation comparison with ground truth on CASIA-v4-Interval.

accuracy, while averaging performance for the "bad" set with deviating information, rather than consistently eliminating over- and undersegmentation errors. Table III illustrates this observation based on E_1 and E_2 error rates comparing segmentation results on both databases for the CAHT and WAHET combinations (for IITD we used dataset-optimised parameters to increase the set of segmentation-consistent images). Given that small segmentation errors are likely to be tolerated by the feature extraction algorithm, we identify the reduction in outliers as a strong factor in the overall improvement, which is unlikely to be reflected in ground-truth-based evaluations aiming to identify statistically significant improvements over the entire set. In the following some of the outliers will be discussed to make the fusion impact clearer.

C. Analysis of Fusion Behaviour

For mask fusion, Fig. 5 shows samples from CASIA v4 interval database. Both the resulting segmentation as well as

(a) Shape mismatch correction. (b) Boundary mismatch correction. (c) Sample discrepancy due to cut off iris. (d) Matching errors.
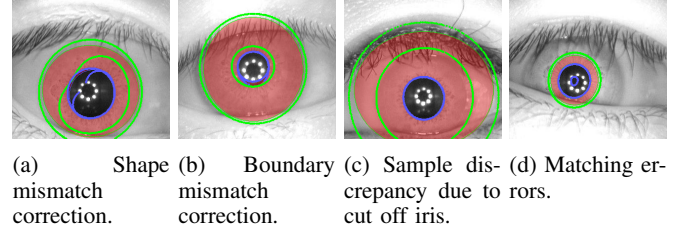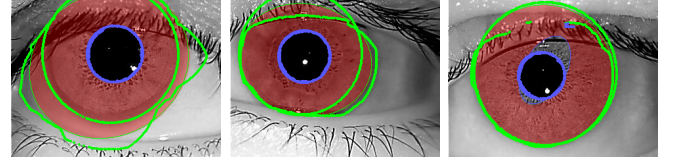
Fig. 5: Possible effects of combining masks.

(a) Detection flaw. (b) Missed boundary. (c) Pruning failure.

Fig. 6: Boundary overestimation and non-convex masks.

the point clouds for iris and pupil boundary are given. The correction behaviour is due to the least-squares ellipse fitting valuing the outer boundaries higher. This leads to corrective behaviour when on of the masks has detected the wrong shape (5a) or the wrong boundary, in this case collarette instead of iris (5b). There are however limits to this, like in case the iris boundary being cut off, leading to a shape bias in the fitting process as seen in Fig. 5c. Further, if both original segmentations exhibit the same type of error the fusion can obviously not correct it, see Fig. 5d.

Fundamentally, the mask fusion values boundary points located farther from the center to a greater extent, e.g. Figs. 5a and 5b. As long as the boundary detection of the iris undershoots rather than overshoots the fusion is auto-corrective. A case where the boundary detection overshoots is the OSIRIS fusion with CAHT on the IITD database. The OSIRIS algorithm frequently overestimates the iris boundary. While this is often corrected by the mask provided by OSIRIS the resulting non-convex and miss shaped masks can lead to fusion problems. Examples comprise the cases of detection flaw and corresponding fusion error (Fig. 6a), missed boundary and an almost correct mask (Fig. 6b), and pruning errors due to a non-convex mask which is not sufficiently removed from correct points (Fig. 6c).

Essentially, as long as the boundary estimation is conservative, i.e. underestimates rather than overestimates, the auto-corrective properties of the mask fusion result in an increased performance. The same properties however will reduce the quality of the mask fusion when boundaries are frequently overestimated. Furthermore, non-convexity of the mask can lead to sample points which are attributed to the wrong boundary. These erroneous samples can be pruned to an extent, but non-convex masks always carry the possibility of a deformed pupillary boundary.

V. CONCLUSION AND FUTURE WORK

This paper analysed multisegmentation fusion using pairwise combinations of CAHT, WAHET, IFPP and OSIRIS iris segmentation algorithms, revealing the autocorrective properties of augmented model fusion on masks in most of the

tested cases (best result 0.64% EER for WAHET+CAHT versus 0.99% EER for CAHT only). Evaluations on ground-truth masks and recognition scores indicated, that ground-truth based evaluations are likely to miss corrective behaviour for outliers, which is critical for the overall task. Detailed error-specific analysis revealed case-specific corrective behaviour, which will be a good starting point for future case-specific fusion approaches. Benefits of multisegmentation in contrast to traditional multialgorithm fusion comprise better normalised source images available for feature-independent storage and the ability to focus on the time-consuming segmentation process, where parallelisation and advanced fusion might be most beneficial. Future work will focus on advanced, sequential approaches taking processing time into account.

REFERENCES

[1] H. Proença and L. Alexandre, "Toward covert iris biometric recognition: Experimental results from the NICE contests," *IEEE Trans. Inf. For. & Sec.*, vol. 7, no. 2, 2012.

[2] A. Ross, R. Jillela, J. Smereka, V. Boddeti, B. Kumar, R. Barnard, X. Hu, P. Pauca, and R. Plemmons, "Matching highly non-ideal ocular images: An information fusion approach," in *Proc. Int'l Conf. on Biometrics (ICB)*, 2012.

[3] F. Alonso-Fernandez and J. Bigun, "Quality factors affecting iris segmentation and matching," in *Proc. Int'l Conf. on Biometrics (ICB)*, 2013.

[4] A. A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of Multibiometrics*. Springer, 2006.

[5] A. Uhl and P. Wild, "Fusion of iris segmentation results," in *Proc. 18th Ib. Congr. on Pattern Recog, (CIARP)*, 2013.

[6] J. Daugman, "How iris recognition works," *IEEE Trans. on Circiuts and Systems for Video Technology*, vol. 14, no. 1, 2004.

[7] H. Hofbauer, C. Rathgeb, A. Uhl, and P. Wild, "Image metric-based biometric comparators: A supplement to feature vector-based hamming distance?" in *Proc. Int'l Conf. Biom. Special Int. Group (BIOSIG)*, 2012.

[8] E. Llano, J. Vargas, M. Garca-Vzquez, L. Fuentes, and A. Ramrez-Acosta, "Cross-sensor iris verification applying robust fused segmentation algorithms," in *Proc. Int'l Conf. on Biometrics (ICB), 2015*, 2015, pp. 1–6.

[9] Y. Sanchez-Gonzalez, Y. Cabrera, and E. Llano, "A comparison of fused segmentation algorithms for iris verification," in *Proc. Ib. Congr. Patt. Rec., (CIARP)*, 2014.

[10] R. P. Wildes, "Iris recognition: an emerging biometric technology," in *Proc. of the IEEE*, vol. 85, 1997.

[11] A. Abhyankar and S. Schuckers, "Active shape models for effective iris segmentation," in *Proc. of SPIE*, 2006.

[12] T. Tan, Z. He, and Z. Sun, "Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition," *Image and Vision Computing*, vol. 28, no. 2, 2010.

[13] Z. He, T. Tan, Z. Sun, and X. Qiu, "Toward accurate and fast iris segmentation for iris biometrics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 9, 2009.

[14] R. Labati, V. Piuri, and F. Scotti, "Agent-based image iris segmentation and multipleviews boundary refining," in *Proc. Int'l Conf. Biom.: Th., Appl. Syst. (BTAS)*, 2009.

[15] G. Sutra, S. Garcia-Salicetti, and B. Dorizzi, "The Viterbi algorithm at different resolutions for enhanced iris segmentation," in *Proc. Int'l Conf. Biom. (ICB)*, 2012.

[16] A. Uhl and P. Wild, "Weighted adaptive hough and ellipsopolar transforms for real-time iris segmentation," in *Proc. Int'l Conf. on Biometrics (ICB)*, 2012.

[17] H. Hofbauer, F. Alonso-Fernandez, P. Wild, J. Bigun, and A. Uhl, "A ground truth for iris segmentation," in *Proc. 22nd Int'l Conf. Pattern Rec. (ICPR)*, 2014.

[18] J. Huang, L. Ma, T. Tan, and Y. Wang, "Learning based resolution enhancement of iris images," in *Proc. BMVC*, 2003.

[19] K. Hollingsworth, T. Peters, K. Bowyer, and P. Flynn, "Iris recognition using signal-level fusion of frames from video," *IEEE Trans. Inf. For. Sec.*, vol. 4, no. 4, 2009.

[20] R. Jillela, A. Ross, and P. Flynn, "Information fusion in low-resolution iris videos using principal components transform," in *IEEE WS Appl. Comp. Vis. (WACV)*, 2011.

[21] Q. McNemar, "Note on the sampling error of the difference between correlated proportions of percentages," *Psychometrika*, vol. 12, no. 2, pp. 153–157, 1947.

[22] A. Fitzgibbon, M. Pilu, and R. B. Fisher, "Direct least square fitting of ellipses," *IEEE Trans. Pat. An. Ma. Int.*, vol. 21, no. 5, 1999.

[23] C. Rathgeb, A. Uhl, and P. Wild, *Iris Recognition: From Segmentation to Template Security*, ser. Advances in Information Security. Springer, 2012, vol. 59.

[24] D. Petrovska and A. Mayoue, "Description and documentation of the biosecure software library," Project No IST-2002-507634 - BioSecure, Tech. Rep., 2007.

[25] A. Uhl and P. Wild, "Multi-stage visible wavelength and near infrared iris segmentation framework," in *Proc. Int'l Conf. Image An. Rec. (ICIAR)*, ser. LNCS, 2012.

[26] "Iris segmentation ground truth database – elliptical/polynomial boundaries (IRISSEG-EP)," http://www.wavelab.at/sources/Hofbauer14b.

[27] L. Ma, T. Tan, Y. Wang, and D. Zhang, "Efficient iris recognition by characterizing key local variations," *IEEE Trans. Image Proc.*, vol. 13, no. 6, 2004.

[28] L. Masek, "Recognition of Human Iris Patterns for Biometric Identification, MSc thesis, Univ. Western Australia, 2003."

[29] "CASIA-IrisV4 Interval Database," http://biometrics.idealtest.org/dbDetailForUser.do?id=4.

[30] "IIT Delhi Iris Database," http://www4.comp.polyu.edu.hk/~csajaykr/IITD/Database_Iris.htm.

[31] A. Edwards, "Note on the "correction for continuity" in testing the significance of the difference between correlated proportions," *Psychometrika*, vol. 13, no. 3, pp. 185–187, 1948.