Quality-based iris segmentation-level fusion

Article

Published Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

CrossMark

Quality-based iris segmentation-level fusion

Peter Wild[1]* , Heinz Hofbauer[2], James Ferryman[3] and Andreas Uhl[2]

Abstract

Iris localisation and segmentation are challenging and critical tasks in iris biometric recognition. Especially in non-cooperative and less ideal environments, their impact on overall system performance has been identified as a major issue. In order to avoid a propagation of system errors along the processing chain, this paper investigates iris fusion at segmentation-level prior to feature extraction and presents a framework for this task. A novel intelligent reference method for iris segmentation-level fusion is presented, which uses a learning-based approach predicting ground truth segmentation performance from quality indicators and model-based fusion to create combined boundaries. The new technique is analysed with regard to its capability to combine segmentation results (pupillary and limbic boundaries) of multiple segmentation algorithms. Results are validated on pairwise combinations of four open source iris segmentation algorithms with regard to the public CASIA and IITD iris databases illustrating the high versatility of the proposed method.

Keywords: Iris biometrics, Segmentation, Fusion, Quality

1 Introduction

Personal recognition from human iris (eye) images comprises several steps: image capture, eye detection, iris localisation, boundary detection, eyelid and noise masking, normalisation, feature extraction, and feature comparison [1]. Among these tasks, it is especially iris localisation and pupillary/limbic boundary detection which challenge existing implementations [2], at least for images captured under less ideal conditions. Examples of undesirable conditions are visible light imaging with weak pupillary boundaries, on-the-move near infrared acquisition with typical motion blur, out-of-focus images, or images with weak limbic contrast.

As an alternative to the development of better individual segmentation algorithms, iris segmentation fusion as a novel fusion scenario [3] was proposed in [4]. For vendor-neutral comparison, this form of fusion has certain advantages over more common multi-algorithm fusion, where each algorithm uses its own segmentation routine: it facilitates data exchange offering access to the normalised texture, increases usability of existing segmentation routines, and allows faster execution requiring only a single module

rather than entire processing chains. In [5], which is extended by this work, a fusion framework for the automated combination of segmentation algorithms is presented, but without taking segmentation quality into account. The reference method in [5] was shown to improve results in many cases, but no systematic improvement could be achieved. A more efficient combination technique can be obtained when inaccurate information can be discarded from the fusion stage, which is the scope of work in this paper. The proposed fusion algorithm assesses the usefulness of individual segmentation input to avoid a deterioration of results even if one of two segmentation results to be combined is inaccurate.

The contributions of this paper are as follows: (1) a generalised fusion framework for combining iris segmentation results extending [5] towards including quality-based predictors of segmentation performance guiding the selection of contributing information (see Fig. 1); (2) a reference implementation based on neural networks and augmented model-based combination of segmentation evidence using iris mask post-processing (such that the only input needed is a segmentation mask file by each algorithm to be considered); and (3) an evaluation of proposed methods analysing pairwise combinations of

*Correspondence: peter.wild@ait.ac.at
[1]AIT Austrian Institute of Technology GmbH, 2444 Seibersdorf, Austria
Full list of author information is available at the end of the article

Wild *et al. EURASIP Journal on Information Security* (2016) 2016:25

Page 2 of 12
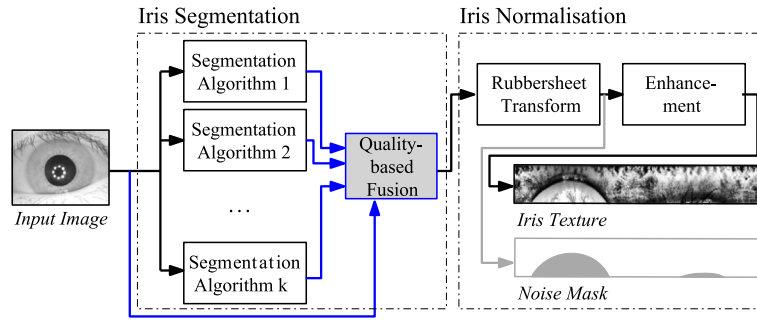
Fig. 1 Proposed framework for fusion of iris segmentation results

algorithms with regard to two aspects: first, conformity with ground truth is inspected focusing on the question whether segmentation fusion concepts indeed improve ground truth accuracy in terms of E_1 and E_2 segmentation errors. Second, the impact on recognition accuracy in terms of receiver operating characteristics (ROCs) and equal error rate (EER) is validated, assuring that segmentation improvement indeed induces less distortions, which in the past has shown to be not necessarily indicated by better ground truth performance [6].

The remainder of the paper is organised as follows. Section 2 presents the methodology and gives an overview of related work in iris fusion, focusing on multi-segmentation, data interoperability, and segmentation quality in iris recognition. The suggested framework and reference method for iris segmentation fusion is presented in detail in Section 3. Section 4 introduces the databases and algorithms under test and gives a detailed presentation and analysis of experiments. Finally, a conclusion of this work and outlook on future topics in segmentation-level iris fusion is given in Section 5.

2 Methodology and related work

The aim of iris segmentation is to retrieve the iris region from an eye image, classifying each pixel location (x, y) into being either out-of-iris or in-iris: $N(x, y) \in \{0, 1\}$. However, given the circular (elliptic, respectively, for out-of-axis acquisitions) shape of the iris, the ultimate outcome needed for iris normalisation is a parameterisation of inner and outer iris boundaries $P, L : [0, 2\pi) \rightarrow [0, m] \times [0, n]$ enclosing non-zero values (iris pixels) in N (ignoring noise and occlusions to avoid non-linear distortions [7]). Using these boundaries, the iris texture is mapped into a coordinate system spanning angle θ and pupil-to-limbic radial distance r [8]. A rubbersheet map $R(\theta, r) := (1 - r) \cdot P(\theta) + r \cdot L(\theta)$ is used to unroll an iris image into a rectangular normalised texture image $T = I \circ R$ (I is the original $n \times m$ image) and normalised noise masks $M = N \circ R$, independent of pupillary

dilation. Often, iris segmentation and normalisation are unified in a single module, however, for fusion purposes it is desirable to separate these two tasks.

Traditional iris segmentation assumes circular boundaries (e.g. [8, 9]). Strong input assumptions often help in case of contradictory information [1] (e.g. very low pupillary contrast, visible light images) and provide especially good performance for cooperative environments [10]. More advanced and relaxed elliptical models (active shape [11], clustering-based [12] Viterbi algorithm [13], weighted adaptive Hough, and ellipsopolar transforms [7]) for P, L provide more accurate segmentation results especially for off-axis images; however, by allowing higher degrees of freedom, they are also prone to errors if preconditions are not met and easily misled. Ideally, advantages of algorithms are combined effectively; however, in this case, some segmentation results have to be rejected, based on the accuracy of the segmentation.

2.1 Segmentation accuracy

Segmentation accuracy is usually computed by analysing noise mask output N with regard to ground truth (mask G) segmentation classification errors, i.e. it is necessary to have manual segmentation ground truth available. Hofbauer et al. [10] collected and released ground truth datasets (IRISSEG-EP) as part of their study. Error rates E_1, E_2 based on classification error are well-employed [14] error measures, differing by whether a priori probabilities are considered (E_2) or not (E_1):

$$E_1 := \frac{1}{k} \sum_{i=1}^{k} \frac{fp_i + fn_i}{mn}; \quad E_2 := \frac{1}{2} \left(\frac{1}{k} \sum_{i=1}^{k} \frac{fp_i}{fp_i + tn_i} \right)$$
$$+ \frac{1}{2} \left(\frac{1}{k} \sum_{i=1}^{k} \frac{fn_i}{fn_i + tp_i} \right),$$

(1)

with tp_i, fp_i denoting the pixel based true and false positive classifications and tn_i, fn_i true and

false negatives for image index i (with dimension $m \times n$).

A problem identified in [6] is that segmentation errors due to low image quality are not necessarily revealed by comparison-based assessment. As cross-comparisons employ the same tool for segmenting the original image and reference template, systematic errors might have a positive overall effect introducing system bias. In order to avoid this bias for the final assessment, we employed ground truth-based segmentation for the reference template to judge for segmentation accuracy testing the sample.

2.2 Iris segmentation quality

As an important aspect of this paper, segmentation performance is evaluated using both segmentation accuracy and impact on recognition performance (evaluating the entire iris processing chain). Alonso-Fernandez et al. [6] have shown how quality indicators and segmentation performance relate to each other; however, they also found that segmentation and recognition performance might be affected by different factors. In this context, it is important to note that systematic segmentation errors can have positive effects, while some segmentation errors might be corrected during pooling stages in feature extraction or rotational alignment at comparison stages. Wei et al. [15] uses defocus, motion blur, and occlusion as iris image quality measures for image selection. ISO/IEC 29794-6 establishes a standard on iris quality. Investigated quality components comprise scalar quality, grey level spread, iris size, pupil-iris ratio, usable iris, iris-sclera contrast, iris-pupil contrast, iris shape, pupil shape, margin, sharpness, motion blur, signal-to-noise ratio, and gaze angle [16]. Wild et al. [17] showed that quality-based filtering can have a pronounced impact on accuracy (up to factor three observed), possibly shadowing potential temporal effects. They also raised the need for transparent recording conditions and pre-evaluation of quality in underlying databases for accurate assessments. In this paper, we use ideas in [17] to develop a ground truth performance predictor as quality indicator for each algorithm to be combined.

2.3 Iris segmentation fusion

Segmentation fusion can be grouped into approaches combining detected boundaries prior to any rubbersheet transformation [4, 5] and after normalisation, where normalised texture is combined [18, 19]. The latter requires multiple execution of the iris unwrapping and normalisation (slower), hiding potential segmentation errors and therefore making their elimination more complex (combination of texture). Most of them implement data-level fusion for superresolution from multiple video frames, such as [20, 21]. State-of-the-art in this context are principal components transform [2, 18] combining multiple normalised iris textures at image-level obtained by different segmentation algorithms. As representative of the first group, Uhl et al. [4] suggested different strategies to combine direct segmentation boundaries rather than texture feeding the combined model into the normalisation routine. Experiments for human (manual) ground truth segmentation showed improved recognition accuracy independent of the employed feature extraction algorithm. The type of fusion method (combination of boundary points for fitting routine vs. interpolating fitted boundaries) did not have a pronounced impact on accuracy. While in [4], for the employed data outliers were not an issue; they were rather critical in [5], where combinations of automated segmentation algorithms did not improve in all cases. Therefore, this work focuses on integrating quality prediction for more efficient fusion at segmentation-level.

The presented work follows the first group. More specifically, in contrast to [18], a single image only is required and unlike [19] normalisation is executed only once. According to our knowledge, this paper is the first to present a quality-driven fusion at segmentation-level in iris recognition. Apart from ISO/IEC TR 24722:2015 (standard on multibiometric fusion, but no support for multinormalisation) and ISO/IEC 19794-6:2011 (segmentation-only *cropped and masked* exchange format), there is no standardisation in segmentation-level fusion.

3 Proposed multi-segmentation fusion method

The proposed multi-segmentation fusion method implementing the framework in Fig. 1 uses noise masks as results of individual segmentation results to generate a best-fitted inner P and outer L boundary curve marking the true possibly occluded iris within the eye image (of course also a corresponding resulting noise mask can easily be generated). It consists of the following four steps, realised as sub-modules, which are explained in detail: (1) *Tracing* derives traced boundaries P_i, L_i for each (ith) candidate through scanning masks and pruning outliers; (2) *Model fusion* combines candidate boundaries (all possible combinations for multiple algorithms); (3) *Prediction* calculates an estimate of the ground truth segmentation error for a particular (combined or individual) segmentation trace P_i, L_i based on a multi layer perceptron assessing quality parameters; and (4) *Fusion selection* acts as a multiplexer returning the (combined or individual) segmentation candidate with the lowest predicted segmentation error. Figure 2 illustrates the process.

3.1 Step 1: tracing

While some iris software offers direct access to iris boundaries P, L after segmentation (e.g. OSIRIS v4.1), other

Fig. 2 Proposed quality-based iris segmentation fusion method
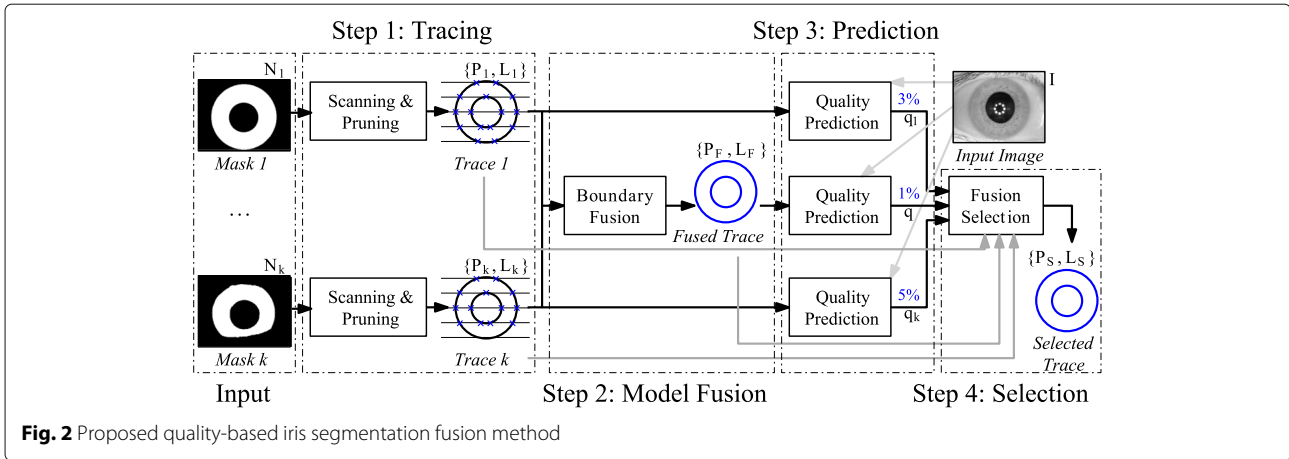
toolkits lack this feature. In addition, there is no unified, but different (e.g. elliptical vs. circular or spline-based) boundary models. Most available models allow for an output of binary noise mask N indicating iris pixels; therefore, we extract P and L from N via the following scanning and pruning process (see Fig. 3).

1. Mesh grid phase: A total of n equidistant scan lines ($n = 100$ yields reasonable results for the employed datasets) are intersected with binary noise mask N locating 0-1 and 1-0 crossings. Based on count and first occurrence, an estimate of limbic or pupillary boundary membership is conducted. Topological inconsistencies (holes) in N should be closed morphologically prior to scanning.
2. Pruning phase: Outlier candidate points with high deviation (radius with z-score ≥ 2.5) from the centre of gravity C_r are removed to avoid inconsistencies in case the outer mask of an iris is not convex, to tolerate noise masks where eyelids are considered and to suppress classification errors.
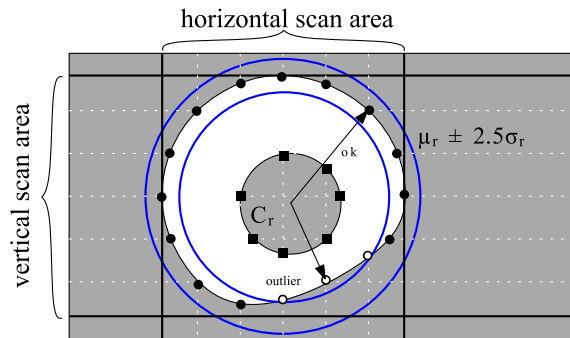
There are some caveats: first, it is not necessarily possible to differentiate between iris and eyelid purely based on the mask—pruning and succeeding model-fitting helps

to reduce such effects. Second, some algorithms employ different boundary models for rubbersheet mapping and noise masks (see [22]). Even in the recent 4.1 version of OSIRIS, noise masks extend over actual boundaries used for unrolling the iris image [5], which has been corrected in experiments by limiting the mask to the employed rubbersheet limbic boundary. Ideally, the employed noise mask for scanning ignores eyelids or other occlusions and a separate noise mask for occlusions is considered at a later stage (e.g. via majority voting after normalisation). While masks may not necessarily be convex and may contain holes, such inconsistencies are repaired by a heuristical algorithm employing simple morphological closing and simplifying local inconsistencies where necessary. Further discussion about this problem and about the method employed can be found in [5].

After scanning and pruning, a set of limbic L_i and pupillary P_i boundary points is available for each (ith) segmentation candidate.

3.2 Step 2: model fusion

Having obtained pupillary and limbic boundary points for each segmentation algorithm, the scope of the model fusion step is to combine a set of segmentation boundaries

Fig. 3 Iris tracing method scanning and pruning iris

Wild *et al. EURASIP Journal on Information Security*   (2016) 2016:25

Page 5 of 12

into a new candidate boundary. This is useful to average individual algorithms's segmentation errors (and considered as "the" fusion method in [5]).

A sequence of k sets of (limbic or pupillary) boundary points $B_1, B_2, \ldots B_k$ can be combined into a single continuous parameterised boundary $B : [0, 2\pi) \rightarrow [0, m] \times [0, n]$ using different strategies, some of which are outlined in Uhl and Wild [4]. The employed fusion technique uses augmented model interpolation [4].

$$\text{Aug rule: } B(\theta) := ModelFit\left(\bigcup_{i=1}^{k} B_i\right)(\theta) \qquad (2)$$

This fusion strategy combines candidate sets $B_1, \ldots, B_k$ into a joint set applying a single parameterisation model *ModelFit* (e.g. least-squares circular fitting) minimising the model-error. This is in contrast to traditional sum rule, where continuous parameterisations are built for each curve to be combined separately. The method is employed separately for inner and outer iris boundaries, and the implementation uses Fitzgibbon's ellipse-fitting [23] for the combination. In this paper, we employ pairwise combinations ($k = 2$); however, the method can easily be extended to test all possible combinations.

It would be a valid choice to use a weighted combination, but we decided against this option, because setting weights requires further tuning with regards to the employed dataset, which we tried to avoid at this stage. The final approach including the following stages uses neural networks which are much more flexible in using features of the image to try and, on a per image basis, weight the fusion.

3.3 Step 3: prediction

With several available (individual and combined) boundary candidates, the critical task of prediction is to obtain estimates on the accuracy of each candidate. Therefore, we employ the set of quality estimators listed in Table 1 to predict segmentation performance, inspired by ISO/IEC 29794-6 the iris quality standard considering device, subject, and environmental impact factors for iris quality estimation. The employed list in this work considers all quality components most recommended by NIST's IREX II iris quality calibration and evaluation [24] (IQCE, a study to examine the effectiveness of ISO/IEC 29794-6 quality components in prediction of performance) that were considered by more than 10 submissions to the evaluation [25]: iris radius (14 submissions), pupil iris ratio (14), iris-sclera contrast (13), iris pupil contrast (13), usable iris area (12), and grey scale utilisation (10). Pupil boundary circularity, and margin adequacy were employed by fewer submissions than the previous metrics and are not included. Ratios were modelled via direct access to radii. Iris pupil concentricity was

Table 1 Quality parameters used for predicting segmentation accuracy

No	Par.	Description	Indicative property
1	p_x	Pupil centre x-coordinate	Close to image centre
2	p_y	Pupil centre y-coordinate	Close to centre
3	p_r	Pupil radius	Sensor-specific distribution (illumination)
4	l_x	Iris centre x-coordinate	Close to centre
5	l_y	Iris centre y-coordinate	Close to centre
6	l_r	Iris radius	System-specific distribution (focus-distance)
7	a_I	Iris area	System/sensor-specific distribution
8	c_P	Pupillary contrast	Usually higher for accurate segmentation
9	c_L	Limbic contrast	Usually higher for accurate segmentation
10	μ	Mean iris intensity	Difficulty of segmentation (eye colour)
11	σ	Iris intensity standard deviation	In-focus assessment (texture)

modelled via direct access to pupil and iris centre positions and included as a recommended measure in [25]. Grey scale utilisation was modelled via mean iris intensity and standard deviation on the iris area only.

Location parameters of circular-fitted pupil and limbic boundary centres (p_x, p_y and l_x, l_y) provide a useful check, whether an iris is found close to image centres (assumed to be more likely for eye patches extracted by preceding eye detectors). Also the distance between centres can potentially reveal segmentation errors. Pupillary and limbic radius values p_r, l_r are included for database-specific predictions. Some segmentation algorithms allow explicit fine-tuning of these segmentation parameters specifying a range of tested values. Successful segmentations are assumed to exhibit sensor (illumination intensity impacting on pupil dilation) or database-specific (focus-distance and average size of human iris) distributions of these parameters. The total available iris texture area with regard to a noise mask (a_I) is an important indicator for noise and challenge of the underlying image. Pupillary and limbic contrast (c_P, c_L) were introduced to judge for the accuracy of fitted boundaries, especially over- and undersegmentation. Boundary contrast is calculated as the absolute difference in average intensity of the circular window (5 pixels height) outside and inside the boundary. Mean iris intensity and standard deviation are included as indicators for potential iris-sclera contrast and focus. All parameters refer to a particular segmentation result (characterised via its noise mask N), more precisely we use the trace (P, L fitted with an elliptical model) after the scanning and pruning stage to compute parameter values.

Wild *et al. EURASIP Journal on Information Security*   (2016) 2016:25

Page 6 of 12

Table 2 Iris datasets used in experiments

Database	Set/ version	Total images	Total classes	Training images	Training classes	Testing images	Testing classes	Band	Resolution
CASIA [27]	v4-Interval	2639	249 (×2)	1320	1-116L	1319	116R-249	NIR	320 × 280
IITD [28, 29]	v1.0	2240	224 (×2)	1120	1-112	1120	113 -224	NIR	320 × 240

When closely looking at characteristics, it is evident to see that multiple effects (overall challenge, accuracy of the segmentation) are present. Therefore, we train a multi layer perceptron to predict E_2 segmentation errors for a segmentation result (P, L) using the $n = 11$ described quality parameters obtained for this result:

- *Input* of the network are the quality parameter values $x_0, \ldots, x_n \in [0, 1]$ obtained from a segmentation result after (min-max) normalisation.
- *Output* of the network is a hypothesis value $h_W(x) = f(W_1^T f(W_2^T x))$ estimating the segmentation error. It is calculated using (trained) matrices W_1, W_2, i.e. a simple neural network with one hidden fully-connected $n \times n$ layer and a logistic-regression output layer. We use $f(z) = \frac{1}{1+\exp(-z)} \in [0, 1]$ (sigmoid) as the activation function.

Matrices W_1, W_2 are trained using m pairs $(x^{(i)}, y^{(i)})$ of correspondences between quality parameters and segmentation ground truth error with regard to a manual human segmentation. The cost function J (using $\lambda = 10^{-6}$) is computed via the limited memory Broyden-Fletcher-Goldfarb-Shanno algorithm [26]:

$$
\begin{aligned}
J(W) = &\frac{1}{m} \sum_{i=0}^{m} \left(\frac{1}{2} \| h_W(x^{(i)}) - y^{(i)} \|^2 \right) \\
&+ \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l-1} \sum_{j=1}^{s_{l+1}-1} \left(W_l[j, i] \right)^2.
\end{aligned}
\tag{3}
$$

We used 50 % of each iris database for training and the remaining 50 % for testing. The computed hypothesis value is the returned quality score $q(P, L) := h_W(x(P, L))$ of a segmentation result.

3.4 Step 4: selection
Note, that the simple combination of boundary curves as suggested and tested in [5] does not always lead to over-all improvements. A substantially inaccurate algorithm can create outliers impacting on overall recognition accuracy. Instead, the proposed method uses predicted quality scores $q(P_i, L_i)$ for each of m candidate segmentations P_i, L_i (individual and combined boundaries using subsets of algorithms) to select the final segmentation index s (and corresponding overall boundary result P_s and L_s) among candidates as follows:

$$
s := \arg \min_{1 \le i \le m} \left(q(P_i, L_i) \right). \tag{4}
$$

Outliers are implicitly removed by considering the combination maximising quality (minimum predicted segmentation error q). Finally, the selected boundary P_s, L_s is used for the rubbersheet transform. Further local noise masks can be combined using e.g., majority voting (not executed). The segmentation tool from [10] is used for unrolling the iris image. It should also be noted that the mask-level fusion generates a mask which is used for unrolling the iris only. 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.

4 Experimental study
We employ the public CASIA [27] and IITD [28, 29] iris databases (see Table 2 for detailed information) in experimental tests. For verification of the positive effect of multi-segmentation fusion, each set is divided into equally-sized disjoint training and test subsets. Training images are needed for learning segmentation accuracy (step 3: prediction) based on the quality indicators introduced in Section 3. All ground truth and recognition accuracy assessments refer to test set images only.

In order to minimise the risk of algorithm-specific impact, we test the proposed framework on four different

Table 3 Segmentation algorithms used in experiments

Algorithm	Implementation	Model	Methods
CAHT [1]	USIT v1.0 [1]	Circular	Circular Hough transform, contrast-enhancement
IFPP [22]	USIT v1.0 [1]	Elliptic	Iterative Fourier approximation; pulling and pushing
OSIRIS [30]	Iris-Osiris v4.1 [30]	Free	Circular HT; active contours
WAHET [7]	USIT v1.0 [1]	Elliptic	Adaptive multi-scale HT; ellipsopolar transform

Wild *et al. EURASIP Journal on Information Security* (2016) 2016:25

Page 7 of 12

Table 4 Feature extraction algorithms used in experiments

Algorithm	Implementation	Length	Comparison	Methods
QSW [32]	USIT 1.0 [1]	10240 bit	Hamming distance	Min-max dyadic wavelet
LG [33]	USIT 1.0 [1]	10240 bit	Hamming distance	1-D Gabor phase quantisation

segmentation algorithms: contrast-adjusted Hough transform (CAHT), weighted adaptive Hough and ellipsopolar transforms (WAHET), iterative Fourier-based pulling and pushing (IFPP), and the open source iris recognition toolkit (OSIRIS) as representatives for elliptic, circular, and free-form iris segmentation models (see Table 3 for an overview listing major methods employed by each technique).

The source code of all segmentation tools is available [1, 30] . Since also the feature extraction technique exhibiting more or less tolerance for segmentation inaccuracies can have an impact on (recognition-based) evaluation results, we test all combinations with two different classical wavelet-based feature extraction techniques: quality assessment and selection of spatial wavelets (QSW) and 1-D Log-Gabor (LG) (see Table 4 for more information). For ground truth segmentation accuracy assessment, we employ the manual segmentations available with [10, 31].

To facilitate reproducible research, the trained neural networks will be made available at http://www.wavelab.at/sources/Wild16a.

4.1 Predictability of segmentation accuracy
We train iris segmentation accuracy prediction separately for each training database, but jointly for all available segmentation algorithms and combinations thereof. Using the true E_1 ground truth segmentation error, we find the minimum (0.189 for CASIA, 0.222 for IITD) of cost function $J(W)$ introduced in Section 3 stopping after 1000 iterations yielding an average delta between prediction and true E_2 error, $\Delta E_2 = 0.017$ for CASIA and $\Delta E_2 = 0.015$ for IITD test sets. This corresponds to 94.03 % accuracy for CASIA and 96.77 % for IITD, respectively, in predicting segmentation errors (considering images with E_2 error > 0.1, i.e. 10 %, as failed segmentations).

From the correlation plots in Figs. 4 and 5, plotting predicted segmentation accuracy $q(P, L)$ versus true $E_2(P, L)$ for each individual algorithm in CASIA and IITD (test sets only), we can see that predictions are quite accurate (note the dual log-scale explaining the wider spread for lower scales) confirming the effectiveness of the proposed simple neural network-based technique.
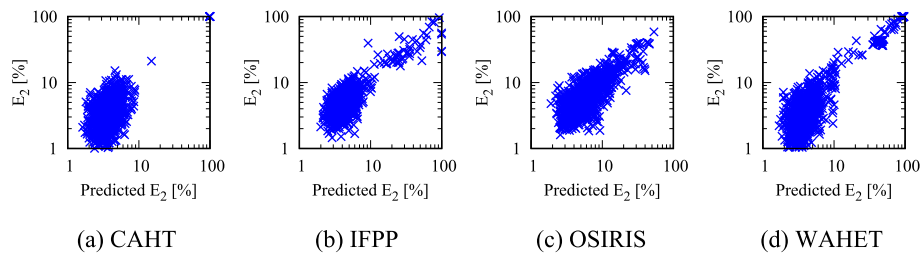
4.2 Ground truth segmentation accuracy
From previous observations in [5], we learned that the combination of segmentation boundaries has the potential to improve "good" segmentation results but may fail producing only averaged results for "bad" segmentation results. Therefore, it is not trivial that with the new quality-based fusion and potential rejection of inaccurate segmentations suggested in this paper indeed better accuracy can be achieved. It has been shown in [5] that there are examples, where simple sum rule fusion degrades overall results compared to the better of the two combined algorithms. However, for some samples and especially when algorithms with tendency for over- and undersegmentation are combined, the combination can be expected to reduce segmentation errors. Figure 6 illustrates a positive example for CASIA (top row, file S1137L06) and IITD (bottom row, file 205-07), where augmented model combination using two segmentation algorithm's outputs (CAHT and WAHET) improves total E_1 (and E_2) segmentation errors. Green areas in the figure indicate false negative iris pixels, while red pixels indicate false positive classifications.

In a second experiment, we evaluated the entire quality-based multi-segmentation fusion method with regard to ground truth segmentation accuracy. Table 5 lists all obtained average E_1 and average E_2 segmentation errors of the test sets for CASIA and IITD iris databases. Individual algorithms are scored at 2.47 % (CAHT), 5.75 %

(a) CAHT (b) IFPP (c) OSIRIS (d) WAHET
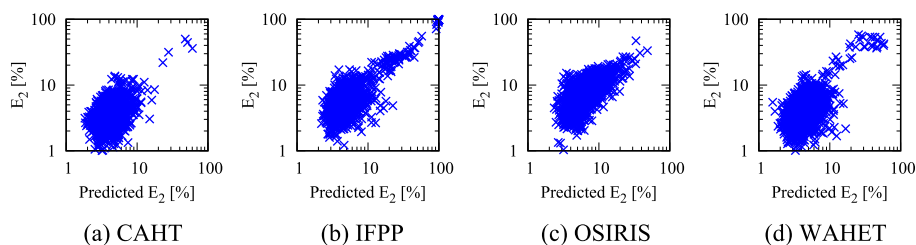
Fig. 4 Predicted versus true segmentation error on CASIA dataset

Wild *et al. EURASIP Journal on Information Security* (2016) 2016:25

Page 8 of 12

Fig. 5 Predicted versus true segmentation error on IITD dataset

(IFPP), 5.27 % (OSIRIS), and 3.45 % (WAHET) E_1 accuracy for CASIA and 2.95 % (CAHT), 4.98 % (IFPP), 5.69 % (OSIRIS), and 5.95 % (WAHET) for IITD. E_2 errors were higher but retained the order of their E_1 counterparts. Of the tested six pairwise combinations using quality-based fusion on two databases, all but a single case returned a better E_1 and E_2 error. In extreme cases, E_1 (and E_2) errors were almost halved (OSIRIS+WAHET 3.10 % vs. OSIRIS 5.69 % and WAHET 5.95 % for IITD). On average, E_1 and E_2 errors were reduced by approx. one tenth (CASIA) to approx. one fifth (IITD) of their original value. However, the amount of reduction varies greatly and certainly depends on whether the two combined algorithms fail for similar images or provide complementary information. Even the case that did not improve results only slightly degraded performance (CAHT+IFPP with 2.56 % E_1 vs. 2.47 % for CAHT on CASIA). This is interestingly the combination with the largest discrepancy in segmentation accuracy for CASIA. A closer look at outlier counts confirmed that the proposed fusion approach could significantly reduce the number of images with a ground truth

segmentation error exceeding 10 % E_1 in all but the mentioned CAHT+IFPP case, where the value stayed the same (tem outliers).

4.3 Impact on recognition accuracy

Note that small segmentation errors are likely to be tolerated by the feature extraction algorithm; therefore, we additionally consider a recognition-based assessment. Since the impact on different feature extractors designed to tolerate slight transformations of the underlying image texture (e.g. varying illumination, head rotation, defocus/blur) is imminent, it is especially interesting to see if there are differences between algorithms. ROCs of all tested scenarios for pairwise combinations plotting individual segmentation algorithm's performance with the combined fusion result are given in Fig. 7 for the IITD database and LG feature extractor, Fig. 8 for IITD and QSW, Fig. 9 for CASIA and LG, and Fig. 10 for CASIA and QSW. Reference ground truth performance rates of these databases and algorithms were (in a given order) 0.09, 0.04, 0.77, and 0.34 % EER.
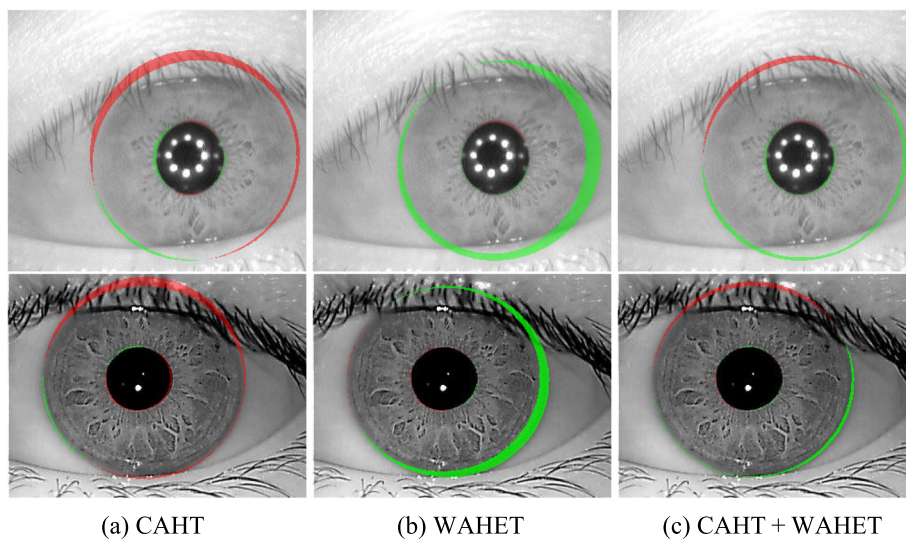
Fig. 6 Sample fusion benefit (*red*: false positives; *green*: false negatives)

Wild *et al. EURASIP Journal on Information Security* (2016) 2016:25
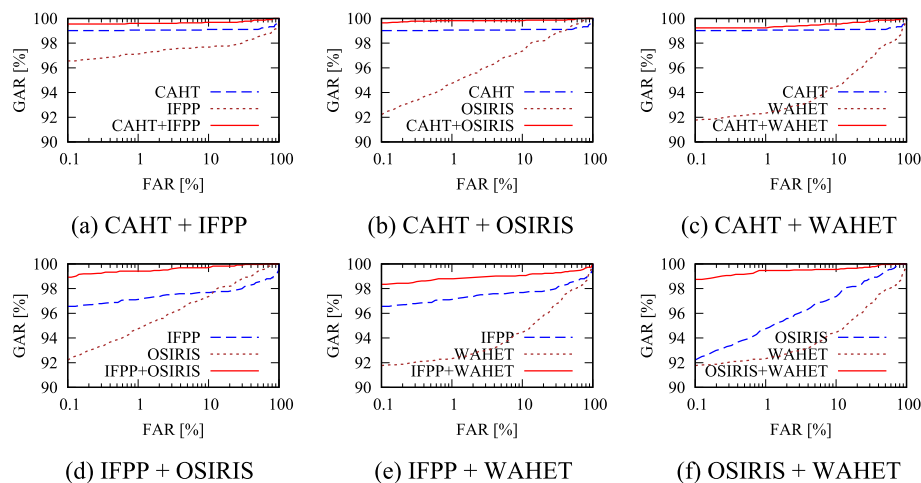
Page 9 of 12

Table 5 Test set segmentation errors (comparison with ground truth) for individual algorithms (diagonal) versus pairwise quality-based segmentation fusion
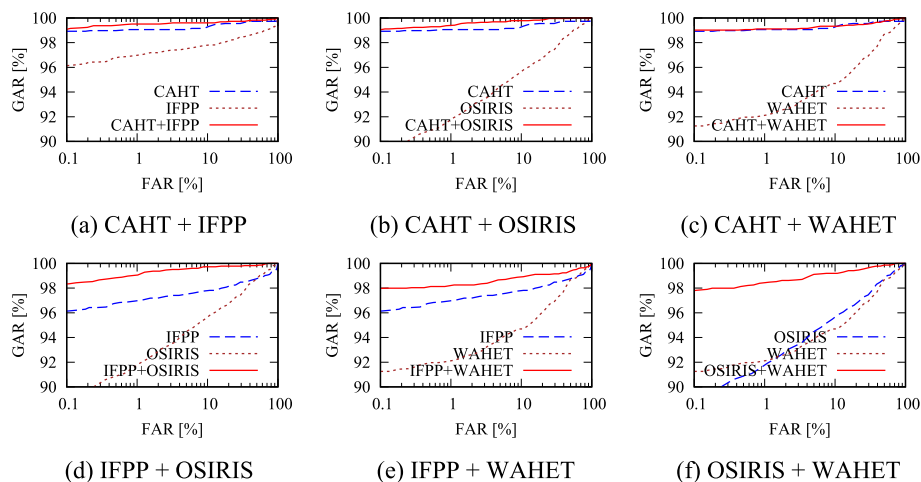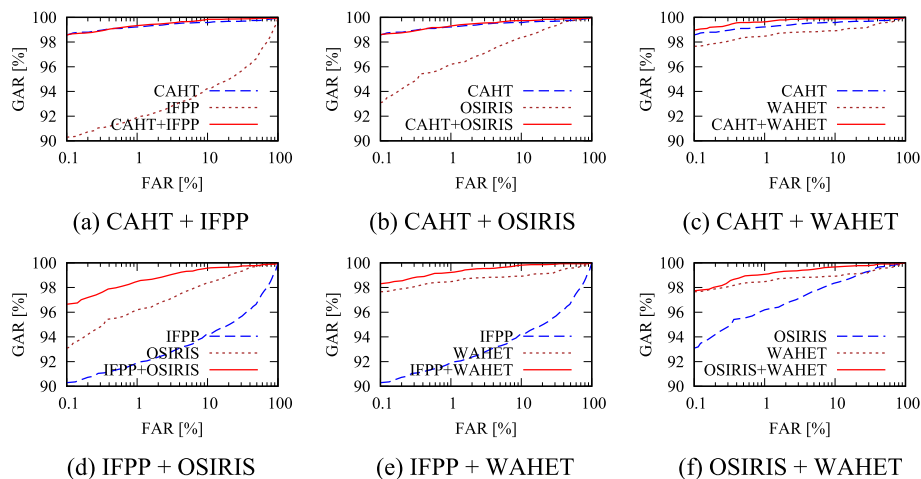
	(a) CASIA v4 interval database				(b) IIT Delhi database			
Average E_1 [%]	CAHT	IFPP	OSIRIS	WAHET	CAHT	IFPP	OSIRIS	WAHET
CAHT	2.47	*2.56*	**2.46**	**2.33**	2.95	**2.88**	**2.73**	**2.80**
IFPP		5.75	**3.67**	**3.08**		4.98	**3.62**	**3.41**
OSIRIS			5.27	**3.01**			5.69	**3.10**
WAHET				3.45				5.95
Average E_2 [%]								
CAHT	3.75	*3.87*	**3.75**	**3.55**	4.28	**3.82**	**3.61**	**3.74**
IFPP		10.09	**5.58**	**4.65**		6.37	**4.77**	**4.49**
OSIRIS			8.20	**4.59**			7.48	**4.11**
WAHET				5.20				7.85
Outlier count ($E_1 > 10$%)								
CAHT	10	*10*	**9**	**6**	10	**6**	**4**	**9**
IFPP		131	**30**	**11**		57	**21**	**24**
OSIRIS			78	**19**			108	**19**
WAHET				38				105

Results where the fused result improves over both individual methods are given in bold, fused results improving over one of the individual methods are given in italics

From Table 6, listing all obtained EERs (rate, where false accepts equals false rejects) for a compact representation of results, it can be clearly seen that now, after feature extraction, quality-based segmentation fusion can improve accuracy in all tested variants, across databases and algorithms. Even in the cases where ground truth segmentation errors were not improved, now recognition rates are slightly better. Single algorithms scored in the range of 1.1–5.97 % EER for CASIA and LG, 0.81–6.48 % for CASIA and QSW, 0.94–6.21 % for IITD and LG, and 0.94–5.86 % for IITD and QSW. Fused algorithms were on average reduced by approx. one fourth (CASIA) to approx.

half (IITD), again depending greatly on algorithms. In some cases, very pronounced improvement was obtained (e.g. OSIRIS+CAHT with 0.25 % EER vs. 0.95 % for CAHT and 3.74 % for OSIRIS in IITD for LG). Especially for algorithms with very weak segmentation accuracies (e.g. OSIRIS 3.74 % and WAHET 6.21 % EER on IITD for LG), remarkable improvements could be achieved when combining the segmentations (OSIRIS+WAHET 0.7 % EER), and improvements were even more pronounced than for the ground truth segmentations. It can therefore be argued that there is additionally a positive effect of the fusion approach to induce less distortions in the

(a) CAHT + IFPP   (b) CAHT + OSIRIS   (c) CAHT + WAHET

(d) IFPP + OSIRIS   (e) IFPP + WAHET   (f) OSIRIS + WAHET

Fig. 7 Quality-based segmentation fusion ROCs on IITD with LG

Fig. 8 Quality-based segmentation fusion ROCs on IITD with QSW
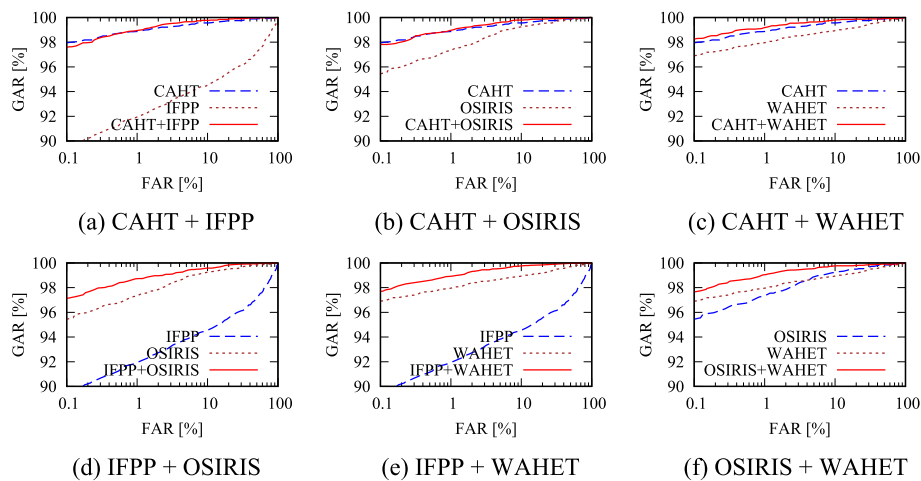
Fig. 9 Quality-based segmentation fusion ROCs on CASIA with LG

Fig. 10 Quality-based segmentation fusion ROCs on CASIA with QSW

Wild *et al. EURASIP Journal on Information Security*   (2016) 2016:25

Page 11 of 12

Table 6 Test set EER performance for individual algorithms (diagonal) versus pairwise quality-based segmentation fusion

	(a) CASIA v4 interval database				(b) IIT Delhi database			
Equal-error rate [%] for LG	CAHT	IFPP	OSIRIS	WAHET	CAHT	IFPP	OSIRIS	WAHET
CAHT	1.10	**1.04**	**1.03**	**0.90**	0.94	**0.45**	**0.25**	**0.76**
IFPP		5.97	**1.24**	**1.06**		2.51	**0.58**	**1.17**
OSIRIS			2.10	**0.96**			3.74	**0.70**
WAHET				1.71				6.21
Equal-error rate [%] for QSW								
CAHT	0.81	**0.74**	**0.79**	**0.43**	0.94	**0.57**	**0.71**	**0.89**
IFPP		6.48	**1.38**	**0.82**		2.66	**0.96**	**1.74**
OSIRIS			2.94	**0.95**			5.41	**1.47**
WAHET				1.38				5.86

Results where the fused result improves over both individual methods are given in bold

mapping phase supporting the overall processing chain. From the ROC figures, it can be seen that improvement is not only pertinent to a specific operating point but largely present across the entire range of operating points. Note, that error rates for the reference technique are different from the rates reported in [5]. In order to use neural networks we had to split the database into a test and trainings set. In order to compare the single with the fused segmentation we had to use the same dataset, the test subset of the database. In [5] the entire database was used for the experiments since the was no training stage.

5   Conclusions

In this paper, we presented a novel quality-based fusion method for the combination of segmentation algorithms. With the positive result for the quality prediction submodule certifying its ability to obtain meaningful estimates of segmentation errors of individual algorithms in as much as 94.03 % (CASIA) to 96.77 % (IITD) of segmentation failures also tests on ground truth segmentation conformity and recognition accuracy confirmed the high versatility of the suggested technique. Analysing pairwise combinations of CAHT, WAHET, IFPP, and OSIRIS iris segmentation algorithms, in all tested cases recognition performance could be improved. The best obtained result for IITD was 0.25 % EER for CAHT+OSIRIS versus 0.94 % EER for CAHT only (using LG) and for CASIA we obtained 0.43 % EER combining CAHT+WAHET versus 0.81 % for CAHT only (using QSW) as best single algorithms. Multi-segmentation fusion has been shown to be a very successful technique to obtain higher accuracy at little additional cost proving to be particularly useful, where better normalised source images are needed. In the future, we will look at even further improved quality prediction and fusion techniques combining multiple segmentation algorithms at once and new sequential approaches saving computational effort. Further, an investigation of extending suggested methods to NIR and VIS images is ongoing work and has shown first promising results. As regards the training of weights, it is clear that also VIS images might be sensitive to different quality metrics and thus should be carefully retrained.

Authors' contributions
PW developed the concept, carried out the quality prediction and ground truth/recognition-based evaluation, and drafted the manuscript. HH provided the mask data, developed the scanning and pruning method, and participated in drafting the manuscript. JF supported in the experimental analysis and revising the manuscript. AU supported in the development of the concept, organised the data, and participated in revising the manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] AIT Austrian Institute of Technology GmbH, 2444 Seibersdorf, Austria. [2] Department of Computer Sciences, University of Salzburg, 5020 Salzburg, Austria. [3] School of Systems Engineering, University of Reading, RG6 6AY Reading, UK.

References
1. C Rathgeb, A Uhl, P Wild, *Iris Recognition: From Segmentation to Template Security*. Advances in Information Security, vol. 59 (Springer, New York, 2012)
2. R Jillela, A Ross, PJ Flynn, in *Proc. Winter Conf. on Appl. Computer Vision, (WACV)*. Information fusion in low-resolution iris videos using principal components transform, (2011), pp. 262–269. doi:10.1109/WACV. 2011.5711512
3. K Ross, AA Nandakumar, AK Jain, *Handbook of Multibiometrics*. (Springer, New York, 2006)
4. A Uhl, P Wild, in *Proc. 18th Ib. Congr. on Pattern Recog, (CIARP)*. Fusion of iris segmentation results, (2013), pp. 310–317. doi:10.1007/978-3-642-41827-3_39

Wild *et al. EURASIP Journal on Information Security*   (2016) 2016:25

Page 12 of 12

5. P Wild, H Hofbauer, J Ferryman, A Uhl, in *Proc. 14th International Conference of the Biometrics Special Interest Group (BIOSIG'15)*. Segmentation-level fusion for iris recognition, (2015), pp. 61–72. doi:10.1109/BIOSIG.2015.7314620

6. F Alonso-Fernandez, J Bigun, in *Proc. Int'l Conf. on Biometrics (ICB)*. Quality factors affecting iris segmentation and matching, (2013). doi:10.1109/ICB.2013.6613016

7. A Uhl, P Wild, in *Proc. Int'l Conf. on Biometrics (ICB)*. Weighted adaptive hough and ellipsopolar transforms for real-time iris segmentation, (2012). doi:10.1109/ICB.2012.6199821

8. J Daugman, How iris recognition works. IEEE Trans. Circuits Syst. Video Technol. **14**(1), 21–30 (2004). doi:10.1109/TCSVT.2003.818350

9. RP Wildes, in *Proc. of the IEEE, vol. 85*. Iris recognition: an emerging biometric technology, (1997)

10. H Hofbauer, F Alonso-Fernandez, P Wild, J Bigun, A Uhl, in *Proc. 22nd Int'l Conf. Pattern Rec. (ICPR)*. A ground truth for iris segmentation, (2014). doi:10.1109/ICPR.2014.101

11. A Abhyankar, S Schuckers, in *Proc. of SPIE*. Active shape models for effective iris segmentation, (2006). doi:10.1117/12.666435

12. T Tan, Z He, Z Sun, Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition. Image Vis. Comput. **28**(2), 223–230 (2010)

13. G Sutra, S Garcia-Salicetti, B Dorizzi, in *Proc. Int'l Conf. Biom. (ICB)*. The Viterbi algorithm at different resolutions for enhanced iris segmentation, (2012). doi:10.1109/ICB.2012.6199825

14. H Proença, L Alexandre, Toward covert iris biometric recognition: experimental results from the NICE contests. IEEE Trans. Inf. For. Sec. **7**(2), 798–808 (2012). doi:10.1109/TIFS.2011.2177659

15. Z Wei, T Tan, Z Sun, J Cui, in *Proc. int'l conf. on biometrics (icb)*, ed. by D Zhang, AK Jain. Robust and Fast Assessment of Iris Image Quality, (2005), pp. 464–471. doi:10.1007/11608288_62

16. E Tabassi, in *Proc. Int'l Conf. Biom. Special Int. Group (BIOSIG)*. Large scale iris image quality evaluation, (2011), pp. 173–184

17. P Wild, J Ferryman, A Uhl, Impact of (segmentation) quality on long vs. short-timespan assessments in iris recognition performance. IET Biometrics. **4**(4), 227–235 (2015). doi:10.1049/iet-bmt.2014.0073

18. E Llano, J Vargas, M García-Vázquez, L Fuentes, A Ramí-rez-Acosta, in *Proc. Int'l Conf. on Biometrics (ICB)*. Cross-sensor iris verification applying robust fused segmentation algorithms, (2015), pp. 17–22. doi:10.1109/ICB.2015.7139042

19. Y Sanchez-Gonzalez, Y Cabrera, E Llano, in *Proc. Ib. Congr. Patt. Rec. (CIARP)*. A comparison of fused segmentation algorithms for iris verification, (2014), pp. 112–119. doi:10.1007/978-3-319-12568-8_14

20. J Huang, L Ma, T Tan, Y Wang, in *Proc. BMVC*. Learning based resolution enhancement of iris images, (2003), pp. 153–162. doi:10.5244/C.17.16

21. K Hollingsworth, T Peters, KW Bowyer, PJ Flynn, Iris recognition using signal-level fusion of frames from video. IEEE Trans. Inf. For. Sec. **4**(4), 837–848 (2009). doi:10.1109/TIFS.2009.2033759

22. A Uhl, P Wild, in *Proc. Int'l Conf. Image An. Rec. (ICIAR)*, ed. by A Campilho, M Kamel. Multi-stage visible wavelength and near infrared iris segmentation framework (2012), pp. 1–10. doi:10.1007/978-3-642-31298-4_1

23. A Fitzgibbon, M Pilu, RB Fisher, Direct least square fitting of ellipses. IEEE Trans. Pat. An. Ma. Int. **21**(5), 476–480 (1999). doi:10.1109/34.765658

24. E Tabassi, PJ Grother, WJ Salamon, *Iris quality calibration and evaluation (IQCE): evaluation report, NIST Interagency/Internal Report (NISTIR), 7820*, (2011), pp. 1–227

25. E Tabassi, in *Proc. of the Int'l Biometrics Performance Conference, IBPC'14*. Iso/iec 29794-6 quantitative standardization of iris image quality, (2014)

26. C Zhu, P Byrd, RH Lu, J Nocedal, Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. ACM Trans. Math. Softw. **23**(4), 550–560 (1997). doi:10.1145/279232.279236

27. CASIA-IrisV4 Interval database. http://biometrics.idealtest.org/dbDetailForUser.do?id=4

28. IIT Delhi iris database. http://www4.comp.polyu.edu.hk/~csajaykr/IITD/Database_Iris.htm

29. A Kumar, A Passi, Pattern Recognition. **43**(3), 1016–1026 (2010). doi:10.1016/j.patcog.2009.08.016

30. D Petrovska, A Mayoue, Description and documentation of the biosecure software library. Technical report, Project No IST-2002-507634 - BioSecure (2007) Available online: http://biosecure.it-sudparis.eu/AB/media/files/BioSecure_Deliverable_D02-2-2_b4.pdf.pdf

31. Iris segmentation ground truth database—elliptical/polynomial boundaries (IRISSEG-EP). http://www.wavelab.at/sources/Hofbauer14b

32. L Ma, T Tan, Y Wang, D Zhang, Efficient iris recognition by characterizing key local variations. IEEE Trans. Image Proc. **13**(6), 739–750 (2004). doi:10.1109/TIP.2004.827237

33. L Masek, Recognition of human iris patterns for biometric identification, MSc thesis, Univ. Western Australia, 2003. Available online: http://www.peterkovesi.com/studentprojects/libor/LiborMasekThesis.pdf