Quality Factors Affecting Iris Segmentation and Matching

Fernando Alonso-Fernandez, Josef Bigun
Halmstad University. Box 823. SE 301-18 Halmstad, Sweden
{feralo,josef.bigun}@hh.se

Abstract

Image degradations can affect the different processing steps of iris recognition systems. With several quality factors proposed for iris images, its specific effect in the segmentation accuracy is often obviated, with most of the efforts focused on its impact in the recognition accuracy. Accordingly, we evaluate the impact of 8 quality measures in the performance of iris segmentation. We use a database acquired with a close-up iris sensor and built-in quality checking process. Despite the latter, we report differences in behavior, with some measures clearly predicting the segmentation performance, while others giving inconclusive results. Recognition experiments with two matchers also show that segmentation and matching performance are not necessarily affected by the same factors. The resilience of one matcher to segmentation inaccuracies also suggest that segmentation errors due to low image quality are not necessarily revealed by the matcher, pointing out the importance of separate evaluation of the segmentation accuracy.

1. Introduction

Iris is rapidly gaining acceptance and support as a viable biometric [17]. In this context, iris image quality assessment is an important trend in the field [17, 12, 5]. Several quality factors can degrade iris images [12]. However, evaluation in the segmentation performance is quite limited [21], with most of the works focused on its impact in the recognition accuracy [12, 17]. Here, we evaluate the impact of 8 quality measures in the performance of iris segmentation. All measures are computed locally (across the iris boundaries) and some of them also globally (in the whole image). We use a segmentation algorithm based on the Generalized Structure Tensor (GST) [1] and the BioSec baseline database (3,200 iris images from 200 contributors in 2 sessions) [8]. Reported results show that in general, local measures are better predictors of the segmentation performance. Different behavior among measures is also observed, with some giving very good discriminative capabilities.

We also evaluate the impact of quality components in the performance of two iris matchers based on Log-Gabor wavelets [16] and SIFT keypoints [3]. The matchers are also observed to be sensitive to quality variations, but not necessarily in the same way than the segmentation algorithm. For instance, with global quality measures, no correlation is found between segmentation and matching performance. Also, the SIFT matcher shows some resilience to segmentation inaccuracies, meaning that errors in the segmentation due to degraded quality may be hidden by the matcher.

The rest of the paper is organized as follows. Sect. 2 describes image properties considered to potentially influence iris recognition accuracy. Section 3 presents the quality measures used. Sect. 4 and 5 describe our experiments and results, respectively, and conclusions are drawn in Sect. 6.

2. Iris image quality

The work [17] defines several image properties considered to potentially influence iris recognition accuracy, in support of development of the standard [4]. This is the first public challenge aimed at identifying algorithm- or camera-agnostic iris image quality components. They include: **Gray scale spread**, with better recognition performance reported with images of high contrast and large dynamic range [17]. **Iris size** (number of pixels across the iris radius, when boundaries are modeled by a circle). **Dilation** (ratio of the pupil to iris radius), with less iris area visible in case of high pupil dilation and higher dissimilarity scores reported in genuine (same person) comparisons between images with different degree of dilation [10]. **Usable iris area** (percentage of non-occluded iris, either by eyelashes, eyelids or reflections). **Contrast of pupil and sclera boundaries**, with sources of variation due to intrinsic (subject character) or extrinsic (illumination or capture device). **Shape (irregularity) of pupil and sclera boundaries**. They are not circular, and not even elliptical, complicating iris segmentation. This irregularity can be natural (anatomical) or due to non-frontal gaze. **Margin** (distance between the iris boundary and the closest image edge). **Sharpness** (absence of defocus blur, which mostly occurs when the focal point is outside the depth of field of the object to be captured). **Motion blur**, caused by the relative

movement of the object and the camera. **Signal to noise ratio**, with the major source believed to be sensor noise. **Gaze angle** (deviation of the optical axis of the eye from the optical axis of the camera, which happens when the subject is not looking directly to the camera). And **interlace** of the acquisition device, caused by two-field CCD sensors. Among these quality components, usable iris area is reported to have the greatest influence on recognition performance, followed by pupil contrast, pupil shape, sclera contrast, gaze angle and sharpness. On the other hand, results for motion blur and signal to noise ratio are inconclusive in [17].

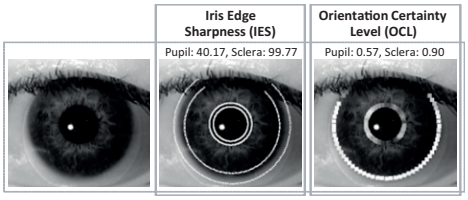


Figure 1. Iris boundary contrast (IES and OCL). Second column shows the points where IES is computed. Third column shows OCL block-wise values across the iris boundaries (brighter color indicate higher quality). Local IES and OCL scores are also given.

3. Computation of iris quality components

In the following, we give details about the quality measures used in this paper. They comprise 8 measures adapted from different algorithms of the literature, or proposed here. We aim to quantify several properties of Section 2. All measures are computed locally (around the pupil or sclera boundaries) and some are also computed in the whole image. Some sample images with different qualities as quantified by the measures used here are shown in Figures 1-3.

(1) Sharpness (defocus blur). This is measured with the iris focus assessment method of [13], which computes the amount of high frequency components. By using a 5×5 convolution kernel, the summated 2-D spectral power is used as focus score. To allow comparison between images or regions of different size, the score is normalized by the actual number of image pixels used.

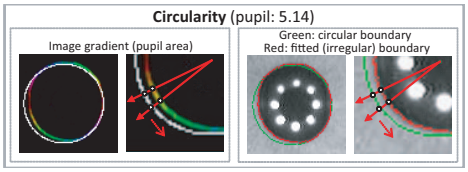


Figure 2. Iris circularity (irregularity). Left: image gradient (the hue encodes the direction and the saturation represents the magnitude). The circle used for boundary modeling is superimposed in white. Right: correspondence in the original image. The circularity score is also given.

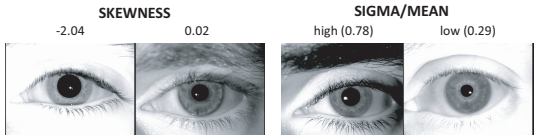


Figure 3. Skewness and standard deviation. First image: negative skewness, most pixels have high gray values. Second image: skewness close to zero, symmetric histogram, no predominance of high or low gray values. Third/fourth images: high/low variability of gray values across the image, respectively.

(2) Motion blur and interlace. These perturbations have the effect of "smearing" the image in the direction of movement. Here, we consider motion blur and interlace to have a similar effect on the image, and we will call it collectively "motion blur". It can be quantified with two parameters [12]: direction (angle) and amount of pixel-smear (strength). As the adjacent rows are quite different in motion blurred images, the difference between every two rows is used as motion blur measurement [19]. A 2×n vertical high-pass spatial filter is used, with the first row with amplitude -1, and the second with 1. We extend this method to account for the direction of motion by rotating the filter with angle increments $\Delta\theta$ and looking for the direction whose filter response has the maximum summated 2-D spectral power. Finally, for size invariance, the score is normalized by the actual number of image pixels used.

(3) Contrast of iris boundaries. This component is quantified with two measures (Figure 1). One is the iris edge sharpness (IES) described in [18]:

$$IES = \sum_{\theta_b \in \zeta} \left(I\left(r_{b+\varepsilon}, \theta_b\right) - I\left(r_{b-\varepsilon}, \theta_b\right) \right) \qquad (1)$$

where $I\left(r, \theta\right)$ is the image intensity in polar coordinates, r_b is the radius of the circle that models the boundary, and θ_b is the angle to move across the circle. IES is computed from two points equidistant to each side of the circle at a distance ε, as can be seen in Figure 1. ζ represents the angles where the boundary is visible. The second measure is based on the Orientation Certainty Level (OCL) proposed in [14] for fingerprints. It measures the energy concentration among the dominant direction of local blocks $W \times W$, computed as the ratio between the two eigenvalues of the covariance matrix of the gradient vector. We use this measure to quantify the strength of the pupil-to-iris and iris-to-sclera boundary transitions. A block of size $W \times W$ is centered at the boundary circle and moved with angle increments $\Delta\theta$. The OCL of all blocks across the boundary is finally averaged.

(4) Circularity of iris boundaries (irregularity). This is computed as follows (Figure 2). We first regularize the iris contours by using radial gradients and active contours (in terms of Fourier series) as in [7]. Given a point (r_b, θ_b) of the circle that models the boundary and the corresponding

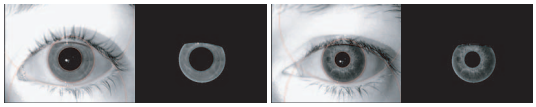


Figure 4. Example of images of the BioSec database with the annotated circles modeling iris boundaries and eyelids.

regularized point (r_m, θ_b), the radial distance $|r_m - r_b|$ is used as circularity measure. This is done for every $\theta_b \in \zeta$, with ζ representing the angles where the boundary is visible. All the distances $|r_m - r_b|$ across ζ are finally averaged. If the boundary is fully circular, the score equals to 0, otherwise it will be higher than 0.

(5) Gray scale spread. This is quantified with two measures: image skewness and standard deviation of gray values (the latter normalized by the mean image gray value, for luminosity independence). The skewness measures the histogram asymmetry. Zero skewness means symmetric histogram. Negative skewness means histogram concentrated to the right (predominance of high gray values), and positive skewness represents the opposite. Examples of these two measures are shown in Figure 3.

(6) Usable boundary (occlusion). This is defined in Section 2 as the percentage of non-occluded iris. For segmentation purposes, we rather use the percentage of non-occluded boundary, when it is modeled by a circle.

4. Iris processing algorithms and databases

We use the iris segmentation algorithm based on the Generalized Structure Tensor (GST) proposed in [1]. The beauty of this method is that, apart from a correlation of edge magnitudes, it takes into account the direction of edges. By using complex filters encoding local orientations of the sought (circular) pattern, its response is penalized if there is disagreement of local orientations of the image with those of the filter. This is not exploited by other edge-based detection methods such as the Circular Hough transform [20] or the Integro-Differential [6] operator, where all boundary pixels contribute equally to (do not penalize) the circle detection. Accordingly, the GST has shown superior performance [1]. This system approximates iris boundaries as circles. Therefore, it outputs the centre/radius of the two boundary circles. Circular detection is the core of most of the literature in iris segmentation [5, 6, 20]. Newer approaches relax the circularity assumption, but many start with a detector of circular edges which are further deformed into non-round boundaries [7, 9, 11].

For recognition experiments, we use two matching algorithms. The first one is the freely available recognition system developed by Libor Masek[1], based on transformation to polar coordinates (using the Daugman's rubber sheet model

[1] www.csse.uwa.edu.au/ pk/studentprojects/libor/sourcecode.html

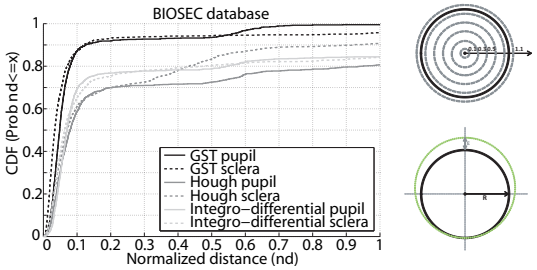


Figure 5. Left: Performance of the automatic segmentation. Top right: relative distance in terms of the radius of the circle. Bottom right: detection accuracy in terms of maximum offset ϵ with respect to the annotated circle. The offset ϵ is normalized by the radius R of the annotated circle for size and dilation invariance.

[6]) and Log-Gabor wavelets plus binary phase quantization [16]. The Hamming distance is used as recognition metric. The second matcher is based on the SIFT operator [15]. We use a free implementation of the SIFT algorithm[2], with the adaptations described in [3]. SIFT keypoints are extracted and matched directly in the original image, without polar coordinates transformation. The recognition metric is the number of matched keypoints between two iris images.

As experimental dataset, we use the BioSec baseline database [8], with 3,200 iris images of 480×640 pixels (height×width) from 200 individuals acquired in 2 sessions with a LG IrisAccess EOU3000 close-up iris camera. Each person contributes with 4 images of the two eyes per session. The EOU3000 sensor has a built-in quality checking process (the best image of a 20 frames video sequence is selected with a proprietary procedure, not disclosed by the manufacturer). Before this sequence acquisition, the camera automatically checks subject' positioning and distance to ensure adequate focus. A set of LED light sources properly positioned ensures that specular reflections fall inside the pupil (Figure 4). We have manually annotated all the images, computing the radius and center of the iris and sclera circles. Similarly, we have also modeled eyelids as circles. Thus, we have also computed the radius and center of those circles. An example of annotated images is shown in Figure 4. In addition, local quality measures are computed around the manually annotated iris boundaries.

5. Results

In Figure 5 (left), we give the segmentation performance of the GST algorithm. We also provide results using the Circular Hough transform (also available in the Libor Masek code) and the Integro-Differential operator (using a public source code[3]). Segmentation accuracy is evaluated in terms of the maximum offset ϵ of the detected circle w.r.t. the annotated one [21]. The offset is normalized by the radius of

[2] http://vision.ucla.edu/ vedaldi/code/sift/assets/sift/index.html

[3] http://web.mit.edu/anirudh/www/

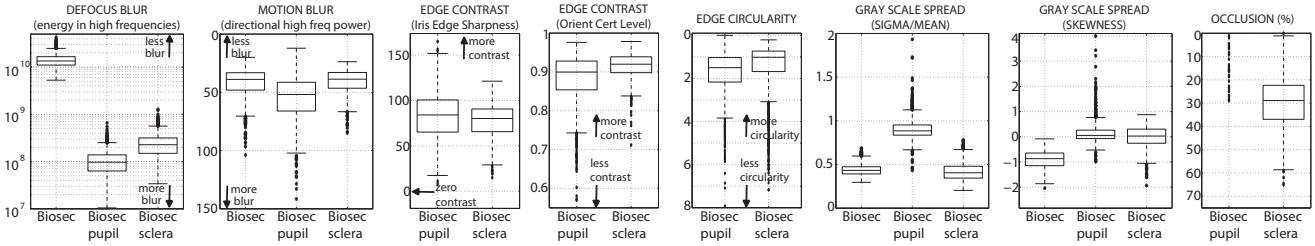
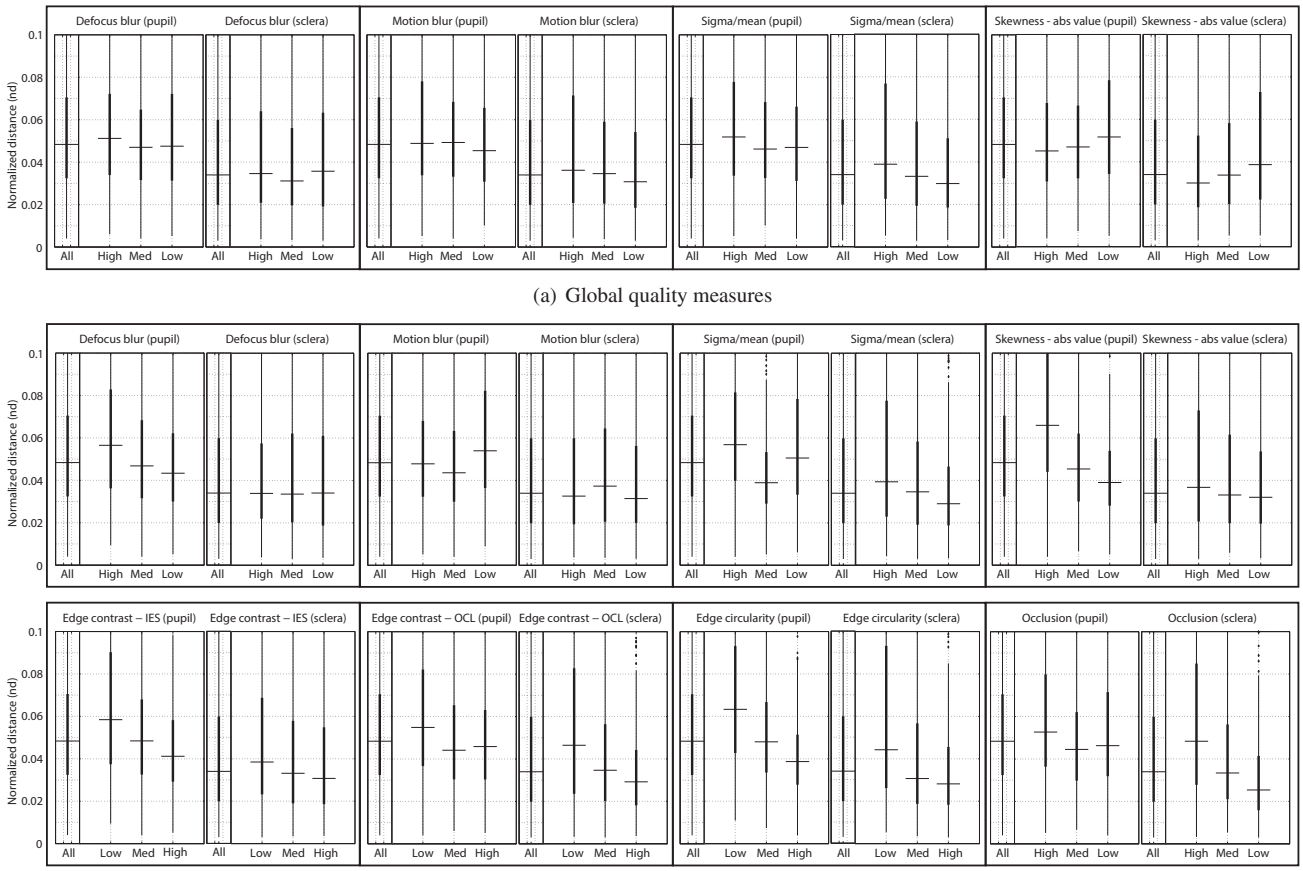


Figure 6. Boxplots of quality measures.



(a) Global quality measures

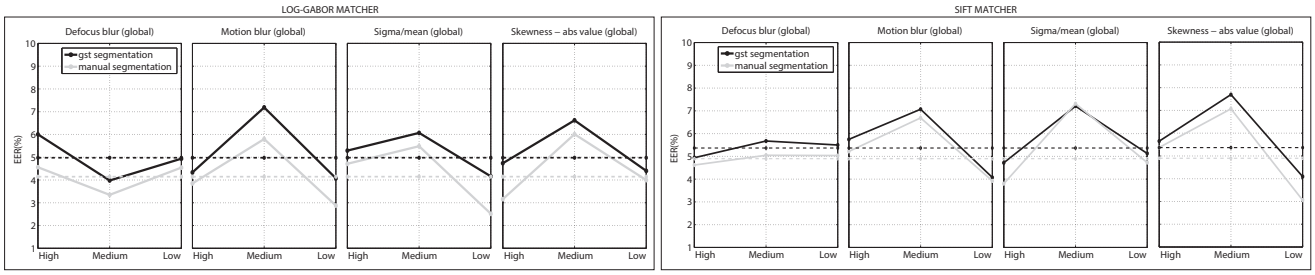
(b) Local quality measures

Figure 7. Performance of the GST segmentation system for quality groups based on the different quality measures used. Performance over the whole database is also given for comparison.

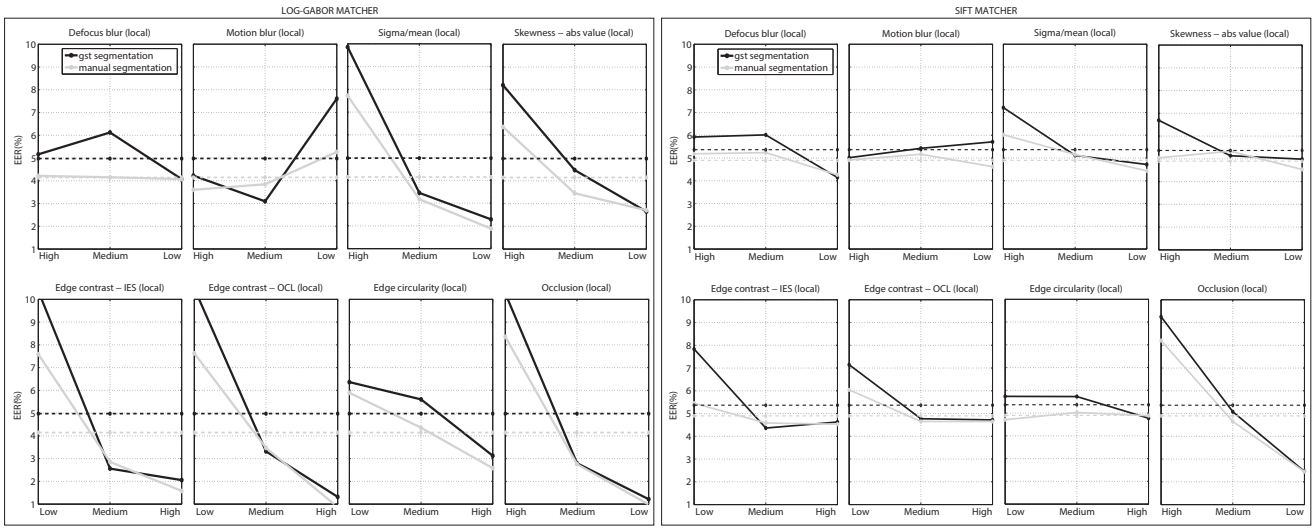
the annotated circle for size and dilation invariance, as illustrated in Figure 5 (right). We observe that the GST algorithm works better than the two other systems, with detected pupil and sclera circles closer to the annotated circles. This superiority has been also observed in previous studies with a different database [1]. For this reason, in the rest of this paper, we only provide results with the GST system.

Figure 6 depicts the distribution of the 8 quality measures used in this paper. To make the defocus blur boxplot more readable, the y-axis is shown in logarithmic scale. Measures *defocus blur*, *motion blur*, *sigma/mean* and *skewness* are computed both locally and globally. An interesting observation is that global and local qualities are not always in the same range. It is worth noting from Figure 6 that the pupil or sclera boundaries exhibit more defocus blur than the whole image, or that the pupil boundary has slightly more motion than the whole image or than the sclera boundary. The latter also happens with the gray scale variability measure (sigma/mean). Also interesting, the gray scale histogram in the pupil and sclera boundaries is highly symmetric (skewness around zero) but in the whole image, it tends to concentrate to higher gray values (negative skewness).

To evaluate the impact of each quality component on the segmentation, we separate all the images of the database



(a) Global quality measures



(b) Local quality measures

Figure 8. EER of the two matchers (Log-Gabor and SIFT) for quality groups based on the different quality measures used. Results are given both with the manual annotation and with the automatic segmentation using the GST. EER over the whole database is also given for comparison (dashed lines).

into three equal-sized quality groups based on each quality measure. Segmentation performance of the GST algorithm for each quality group is then reported (Figure 7). It should be noted that since global and local quality measures are not in the same range, quality groups with the same label (e.g. 'low') do not have to the same range of values. The same applies to the pupil and sclera qualities. For the *skewness* measure, we use its absolute value, so we compare the effect of using samples with zero skewness (symmetric histogram) vs. high absolute skewness (non-symmetric histogram). It can be observed that in general, local quality metrics are better predictors of the segmentation performance than global metrics, which means than better performance can be achieved (compare y-axis ranges of the boxplots). Unfortunately, local quality measures have the obvious limitation of requiring segmentation [12]. Although manual annotation is unfeasible in operational environments, our purpose however is to reveal the sources of error with the aim of guiding subsequent developments and improvements of iris algorithms. It is also worth noting the different behavior of global and local quality measures in some cases (compare the different tendencies of *motion blur* or *skewness*). If we focus only on the local measures, the best predictors of pupil segmentation accuracy are (in this order): i) *skewness*, *sigma/mean* and *circularity*; ii) *edge contrast (IES)*; iii) *edge contrast (OCL)*; and iv) *defocus*. As for the sclera, we have: i) *occlusion*; ii) *sigma/mean*, *circularity* and *edge contrast (OCL)*; and iii) *edge contrast (IES)*. Both pupil and sclera are sensitive to the (low) circularity and (low) edge contrast. This is obvious, since the GST segmentation algorithm is a circular detector and it is based in edge analysis. On the other hand, measures such as *defocus blur*, *motion blur* or *sigma/mean* do not affect equally to the pupil and to the sclera.

Lastly, we evaluate the performance of the two matchers of Section 4 also by partitioning the data into equal-sized quality groups. In this case, the quality of a score is defined as $(Qe+Qt)/2$, where Qe and Qt are the image qualities of the enrolled and input iris respectively corresponding to the matching. When quality is computed locally, Qe, Qt are computed as the average of the pupil and sclera qualities. Results of this procedure are given in Figure 8. We include

performance curves both with the manual annotation and with the automatic segmentation using the GST. An important observation is that the EER using manual annotation (solid gray curves) varies among the quality groups. Assuming that manual segmentation is of high accuracy, this indicates that the matchers are also sensitive to variations in quality. This variability with manual annotation is more evident with the Log-Gabor matcher. In other words, the SIFT matcher is not so sensitive to the quality factors studied.

No correlation is observed between segmentation performance and EER values when quality groups are generated with the global quality measures (compare tendencies between Figures 7 and 8). This suggests that the matchers are affected in the opposite way than the segmentation algorithm. On the contrary, the tendency observed in the segmentation performance of local quality measures is mirrored in the EER nearly in all cases. However, it is worth noting that the SIFT matcher is less sensitive to variations in local quality. This has one positive effect: when segmentation accuracy is bad, the matching performance is not worsened as much as than with the Log-Gabor matcher. But the opposite also occurs: when segmentation accuracy is good, matching accuracy does not improve too much either. An exception is the occlusion measure. It is expected that as the amount of iris texture information is reduced, the performance of the two matchers worsen accordingly, as well as the opposite.

6. Conclusions

The impact of several image quality components in the performance of iris segmentation is evaluated. Quality measures are computed locally (around the iris boundaries) and some of them are also computed globally (in the whole image). It has been found that local quality metrics are better predictors of the segmentation accuracy than global metrics, despite the obvious limitation of requiring segmentation. Some measures also behave differently when they are computed locally or globally.

We also evaluate the impact of quality components in the performance of two iris matchers based on Log-Gabor wavelets and SIFT keypoints. We observe that the matchers are also sensitive to quality variations, but not necessarily in the same way than the segmentation algorithm. Also, the SIFT matcher is observed to be more resilient to segmentation inaccuracies. In this sense, errors in the segmentation may be hidden by the matcher, pointing out the importance of evaluating also the precision of iris segmentation, rather than focusing on recognition accuracy only [21].

Some other preliminary experiments (not given) show that quality measures are not necessarily correlated. Quality is intrinsically multi-dimensional and it is affected by factors of very different nature [2]. Future work includes fusing the estimated quality measures to obtain a single measure with higher prediction capability of the segmentation and matching accuracy [12]. Another source of work will be the different sensitivity observed in the two matchers. By using adaptive quality fusion schemes, we will seek to obtain better performance over a wide range of qualities [2].

Acknowledgments

Author F. A.-F. thanks the Swedish Research Council and the EU for for funding his postdoctoral research. Authors also acknowledge the CAISR research program of the Swedish Knowledge Foundation, the EU BBfor2 project (FP7-ITN-238803) and the EU COST Action IC1106 for its support. Authors also would like to thank L.M. Tato-Pazo for her valuable work in annotating the iris database, and to the Biometric Recognition Group (ATVS-UAM) for making the iris part of the BioSec database available for our experiments.

References

[1] F. Alonso-Fernandez, J. Bigun. Iris boundaries segmentation using the generalized structure tensor. *Proc. BTAS*, 2012.

[2] F. Alonso-Fernandez, J. Fierrez, J. Ortega-Garcia. Quality measures in biometric systems. *IEEE Security and Privacy*, 10:52–62, 2012.

[3] F. Alonso-Fernandez, P. Tome-Gonzalez, V. Ruiz-Albacete, J. Ortega-Garcia. Iris recognition based on sift features. *Proc. IEEE BIDS*, 2009.

[4] D. Benini et al. ISO/IEC 29794-6 Biometric Sample Quality - part 6: Iris image. *JTC1/SC37/Working Group 3 - http://isotc.iso.org/isotcportal*, 2012.

[5] K. Bowyer, K. Hollingsworth, P. Flynn. Image understanding for iris biometrics: a survey. *Computer Vision and Image Understanding*, 110:281–307, 2007.

[6] J. Daugman. How iris recognition works. *IEEE TCSVT*, 14:21–30, 2004.

[7] J. Daugman. New methods in iris recogn. *IEEE TSMC-B*, 37:1167–1175, 2007.

[8] J. Fierrez et al. BioSec baseline corpus: A multimodal biometric database. *Patt. Recogn.*, 40:1389–1392, 2007.

[9] Z. He, T. Tan, Z. Sun, X. Qiu. Toward accurate and fast iris segmentation for iris biometrics. *IEEE TPAMI*, 31:1295–1307, 2010.

[10] K. Hollingsworth, K. Bowyer, P. Flynn. Pupil dilation degrades iris biometric performance. *Computer Vision and Image Understanding*, 113:150–157, 2009.

[11] D. Jeong et al. A new iris segmentation method for non-ideal iris images. *Image and Vision Computing*, 28:254–260, 2010.

[12] N. D. Kalka, J. Zuo, N. A. Schmid, B. Cukic. Estimating and Fusing Quality Factors for Iris Biometric Images. *IEEE TSMC-A*, 40:509–524, 2010.

[13] B. Kang, K. Park. Real-time image restoration for iris recognition systems. *IEEE TSMC-B*, 37:1555–1566, Dec. 2007.

[14] E. Lim, X. Jiang, W. Yau. Fingerprint quality and validity analysis. *Proc. ICIP*, 1:469–472, 2002.

[15] D. Lowe. Distinctive image features from scale-invariant key points. *International Journal of Computer Vision*, 60:91–110, 2004.

[16] L. Masek, P. Kovesi. Matlab source code for a biometric identification system based on iris patterns. *School of Computer Science and Software Engineering, University of Western Australia*, 2003.

[17] E. Tabassi, P. Grother, W. Salamon. IREX II - IQCE - iris quality calibration and evaluation. Performance of iris image quality assessment algorithms. *NISTIR 7296 - http://iris.nist.gov/irex/*, 2011.

[18] Z. Wei, X. Qiu, Z. Sun, T. Tan. Counterfeit iris detection based on texture analysis. *Proc. ICPR*, pp. 1–4, 2008.

[19] Z. Wei, T. Tan, Z. Sun, J. Cui. Robust and fast assessment of iris image quality. *Proc. ICB*, pp. 464–471, 2006.

[20] R. P. Wildes. Iris recognition: An emerging biometric technology. *Proc. IEEE*, 85:1348–1363, 1997.

[21] J. Zuo, N. Schmid. An automatic algorithm for evaluating the precision of iris segmentation. *Proc. IEEE BTAS*, 2008.