This is the submitted version of a paper presented at *IWBF 2014 – 2nd International Workshop on Biometrics and Forensics 2014, Valletta, Malta, 27-28th March, 2014*.

Citation for the original published paper:

Eye Detection by Complex Filtering for Periocular Recognition.  
https://doi.org/10.1109/IWBF.2014.6914250

N.B. When citing this work, cite the original published paper.

Permanent link to this version:  
http://urn.kb.se/resolve?urn=urn:nbn:se:hh:diva-24627
ABSTRACT
We present a novel system to localize the eye position based on symmetry filters. By using a 2D separable filter tuned to detect circular symmetries, detection is done with a few 1D convolutions. The detected eye center is used as input to our periocular algorithm based on retinotopic sampling grids and Gabor analysis of the local power spectrum. This setup is evaluated with two databases of iris data, one acquired with a close-up NIR camera, and another in visible light with a webcam. The periocular system shows high resilience to inaccuracies in the position of the detected eye center. The density of the sampling grid can also be reduced without sacrificing too much accuracy, allowing additional computational savings. We also evaluate an iris texture matcher based on 1D Log-Gabor wavelets. Despite the poorer performance of the iris matcher with the webcam database, its fusion with the periocular system results in improved performance.

Index Terms— Biometrics, periocular, eye detection, symmetry filters

1. INTRODUCTION
Periocular recognition has gained attention recently in the biometrics field [1, 2, 3, 4] due to demands for increased robustness of face or iris systems. It refers to the face region in the vicinity of the eye, including the eye, eyelids, lashes and eyebrows. This region can be easily obtained with existing setups for face and iris, and the requirement of user cooperation can be relaxed. An evident advantage is its availability over a wide range of distances even when the iris texture cannot be reliably obtained (low resolution) or under partial face occlusion (close distances). Most face systems use a holistic approach, requiring a full face image, so the performance is negatively affected in case of occlusion [3]. Also, the periocular region appears in iris images, so fusion with the iris texture has potential to improve the overall recognition [5]. Periocular region has also shown superior performance than face under extreme values of blur or down-sampling [6]. This points out the strength of periocular recognition when only partial face images are available, for example forensics or surveillance, or in more relaxed scenarios such as distant acquisition or mobile devices.

In previous research [7], we proposed a periocular system based on retinotopic sampling grids positioned in the pupil center, followed by Gabor decomposition at different frequencies and orientations. This system, which we evaluated on two databases with near-infrared (NIR) illumination, achieved competitive verification rates in comparison with existing periocular approaches [3, 4]. As in most studies on periocular recognition, it also relied on manual marking of the periocular region. Some recent works have started to deal with the issue of locating the eye position without relying on a full-face detector. The work in [8] used the OpenCV implementation of Viola-Jones detectors of face sub-parts [9]. An accuracy of 96.4/99.2% in the combined detection of face parts was reported using NIR/visible face images, respectively. Eye detection can be also a decisive pre-processing task to ensure successful segmentation of the iris texture in difficult images. In [10], a correlation filter [11] trained with 1000 images was used as eye center detector, achieving a 95% success rate with NIR images from subjects walking through a portal. The output of the eye center detector was used as input to some traditional iris segmentation algorithms, with a segmentation accuracy ranging from 51% to 90% using such difficult images.

Here, we propose a system for eye localization based on symmetry filters. An advantage of our system with respect to the mentioned eye detectors is that it does not need training. Also, by using 2D separable filters, detection can be done quite fast by few 1D convolutions. This detection system is evaluated with two databases of iris data, one acquired with a close-up NIR camera, and another in visible light with a webcam. The proposed detection system is able to give the center of the eye with good accuracy in the NIR database, and still provides good accuracy in the webcam database. This algorithm is validated by using the detected eye center as input to our periocular recognition system, which is evaluated both in verification and identification mode. Results show that the periocular system is quite robust to inaccuracies in detecting the center of the eye. We also evaluate an iris texture matcher based on 1D Log-Gabor wavelets. The performance of this matcher is considerably worse under webcam data, but it is able to complement our periocular system, with improved performance observed by the fusion of the two systems.
ORIENTATION is the complex-valued orientation tensor field given by the computation of the second order complex moment of the symmetry derivatives of Gaussians. The Symmetry filters are a family of filters computed from symmetry filters, Equation 1, which output how much of a certain symmetry exist in a local image neighborhood [12, 13].

**Symmetry filters**

Symmetry filters are a family of filters computed from symmetry derivatives of Gaussians. The \( n \)-th symmetry derivative of a Gaussian, \( \Gamma^{\{n, \sigma^2\}} \), is obtained by applying the partial derivative operator \( D_x + i D_y = (\partial/\partial x) + i (\partial/\partial y) \), called 1\(^{st} \) symmetry derivative, to a Gaussian:

\[
\Gamma^{\{n, \sigma^2\}} = \begin{cases} 
(D_x + i D_y)^n g(x, y) & (n \geq 0) \\
(D_x - i D_y)^n g(x, y) & (n < 0) 
\end{cases} \tag{1}
\]

Since \( D_x + i D_y \) and \( (\frac{1}{\sigma^2}) (x + iy) \) behave identically when acting on a Gaussian [12, 13], Eq. 1 can be rewritten as:

\[
\Gamma^{\{n, \sigma^2\}} = \begin{cases} 
(\frac{1}{\sigma^2})^n (x + iy)^n g(x, y) & (n \geq 0) \\
(\frac{1}{\sigma^2})^n (x - iy)^n g(x, y) & (n < 0) 
\end{cases} \tag{2}
\]

The interest is that these symmetry derivatives of Gaussians are able to detect patterns as those of Figure 1 through the computation of the second order complex moment of the power spectrum via \( I_{20} = \langle \Gamma^{\{n, \sigma^2\}}, h \rangle \), where \( h \) is the complex-valued orientation tensor field given by \( h = \langle \Gamma^{\{1, \sigma^2\}}, f \rangle^2 \) and \( f \) is the image under analysis [13].

For each family of symmetric patterns, there is a symmetry filter \( \Gamma^{\{n, \sigma^2\}} \) (indexed by \( n \)) suitable to detect the whole family [14]. The local maxima in \( I_{20} \) gives the location, whereas the argument of \( I_{20} \) at maxima locations gives the group orientation of the detected pattern (except for the family of column 3 in Figure 1, where the ‘orientation’ represents the chirality of the curves). Therefore, \( I_{20} \) encodes how much of a certain type of symmetry exists in a local neighborhood of the image \( f \). In addition, a single symmetry filter \( \Gamma^{\{n, \sigma^2\}} \) is used for the recognition of the entire family of patterns, regardless of their orientation (or chirality). Symmetry filters have been successfully applied to a wide range of detection tasks such as cross-markers in vehicle crash tests [15], core-points and minutiae in fingerprints [16, 17], or iris boundaries [18]. The beauty of this method is even more emphasized by the fact that \( I_{20} \) is computed by filtering in Cartesian coordinates without the need of transformation to curvilinear coordinates (which is implicitly encoded in the filter).

**Eye detection process**

We use the filter of order \( n = -2 \) to detect the position of the eye in a given image. By assuming that iris boundaries can be approximated as circles, the eye can be detected with the pattern of concentric circles shown in Figure 1 (top left). Despite the inner (pupil) and outer (sclera) boundaries of the iris are not concentric, we exploit the evidence that the pupil is fully contained within the sclera boundary, with the center of both circles in close vicinity [19]. Due to the separable property of 2D gaussians, the filter can be re-written as:

\[
\Gamma^{\{-2, \sigma^2\}} = \left( \frac{1}{\sigma^2} \right)^2 (x - iy)^2 g(x) g(y) \tag{3}
\]

so the 2D convolutions can be computed by several 1D convolutions, achieving a considerable higher speed. Also, in computing \( h \), 1D convolutions can also be used [16]. After the computation of \( I_{20} \), we search for local maxima in \( |I_{20}| \) with a window of size \( 7 \times 7 \). The maximum with highest magnitude is selected as the center of the eye. Evidence of the pattern of concentric circles is given by an argument of the complex filter response equal to zero. Thus, only maxima with absolute angle below a certain threshold are considered.

An example of the detection process is shown in Figure 2.

**3. PERIOcular recognition SYSTEM**

For periorcular recognition, we use the system proposed in [7], which is based on [20, 2]. It makes use of a sparse retinotopic sampling grid positioned in the eye center. The grid has rectangular geometry, with sampling points distributed uniformly (Figure 3). The local power spectrum of the image is sampled at each cell of the grid by a set of modified Gabor filters organized in 5 frequency channels and 6 equally spaced orientation channels. The sparseness of the sampling grid allows direct Gabor filtering in the image domain without needing the Fourier transform, with significant computational savings [2] and even feasibility in real time [21]. The Gabor responses
are grouped into a single complex vector, which is used as identity model. Matching between two images is using the magnitude of complex values. Prior to matching with magnitude vectors, they are normalized to a probability distribution (PDF), and matching is done using the $\chi^2$ distance [22]. Rotation is accounted for by shifting the grid of the query image in counter- and clock-wise directions, and selecting the lowest distance, which corresponds to the best match between two templates.

**4. DATABASE AND BASELINE IRIS MATCHER**

As experimental dataset, we use the BioSec baseline database [23] and the MobBIO database [24]. BioSec has 3,200 iris images of 480×640 pixels (height×width) from 200 individuals acquired in 2 sessions with a LG IrisAccess EOU3000 close-up infrared iris camera. Each person contributes with 4 images of the two eyes per session. MobBIO, on the other hand, has been captured with the Asus Eee Pad Transformer TE300T Tablet (i.e. a webcam in visible light) in one single session. Images were captured in two different lighting conditions, with variable eye orientations and occlusion levels, resulting in a large variability of acquisition conditions. Distance to the camera was kept constant, however. Here, we use the training dataset, which comprises 800 iris images of 200×240 pixels from 100 individuals, with 4 images of the two eyes per person. We have manually annotated the two databases, computing the radius and center of the pupil and sclera circles. Similarly, we have also modeled eyelids as circles, computing the radius and center of those circles too (in BioSec, eyelids groundtruth is available for 75 individuals only). An example of annotated images is shown in Figure 4.

We conduct matching experiments of iris texture using 1D log-Gabor filters [25]. The iris region is unwrapped to a normalized rectangle using the Daugman’s rubber sheet model [19] and next, a 1D Log-Gabor wavelet is applied plus phase binary quantization to 4 levels. Matching between binary vectors is done using the normalized Hamming distance [19], which incorporates the noise mask, so only significant bits are used in computing the Hamming distance. Rotation is accounted for by shifting the grid of the query image in counter- and clock-wise directions, and selecting the lowest distance, which corresponds to the best match between two templates.

**5. EXPERIMENTAL PROTOCOL**

We conduct both verification and identification experiments. We consider each eye as a different user, resulting in 200 available users in MobBIO. For fair comparison with the iris matcher, we use in the recognition experiments only the subset of BioSec with both pupil/sclera and eyelids groundtruth available (75 individuals, or 150 different users). This is to avoid introducing non-significant bits in the noise mask. Verification performance experiments with MobBIO are as follows. Genuine matches are obtained by comparing each image of a user to the remaining images of the same user, avoiding symmetric matches. Impostor matches are obtained by comparing the $1^{st}$ image of a user to the $2^{nd}$ image of the remaining users. We obtain $200 \times 6=1,200$ genuine and $200 \times 199=39,800$ impostor scores. With BioSec, genuine matches for a given user are obtained by comparing the 4 images of the $1^{st}$ session to the 4 images of the $2^{nd}$ session. Impostor matches are obtained by comparing the $2^{nd}$ image of the $1^{st}$ session of a user to the $2^{nd}$ image of the $2^{nd}$ session of all the remaining users. With this, we obtain $150 \times 4 \times 4=2,400$ genuine and $150 \times 149=22,359$ impostor scores. Note that experiments with the BioSec database are
made by matching images of different sessions, but these inter-session experiments are not possible with MobBIO.

In the identification experiments, we define one sample of each eye as enrolment sample, and the remaining samples of the database are used for testing. Given a test image, identification experiments are done by outputting the N closest identities of the enrolment set. An identification is considered successful if the correct identity is among the N outputted ones. With MobBIO, this results in 200×200×3=120,000 computed distances, and 150×150×4=90,000 with BioSec.

Some fusion experiments are also done between the periocular and the iris matchers. The fused distance is computed as the mean value of the distances due to the individual matchers, which are first normalized to be similarity scores in the [0,1] range using tanh-estimators as \[ s' = \frac{1}{2} \left\{ \tanh \left( 0.01 \left( \frac{s - \mu_s}{\sigma_s} \right) \right) + 1 \right\}. \]
Here, s is the raw similarity score, \( s' \) denotes the normalized similarity score, and \( \mu_s \) and \( \sigma_s \) are respectively the estimated mean and standard deviation of the genuine score distribution [26].

### 6. Results

In Figure 5, we give the performance of our eye detection system on the two databases. The symmetry filter used with BioSec has a size of 361×361, and 151×151 with MobBIO. The filter covers approximately 75% of the shortest image side. This is to ensure that it captures the different size of the eyes present in the databases due to small variations in the distance to the sensor. Detection accuracy is evaluated by the distance of the detected eye center with respect to the annotated pupil and sclera centers [27]. Distances are normalized by the radius of the annotated circles for size and dilation invariance, as shown in the inner sub-figure of Figure 5. This way, a normalized distance \( nd \) lower than 1 means that the detected point is inside the circle, and the opposite if \( nd > 1 \). Also, since the sclera radius is always lower than the pupil radius, the normalized distance w.r.t. the sclera center will be smaller than w.r.t. the pupil center, as can be observed in Figure 5.

Considering the database acquired under good, controlled conditions (BioSec), we observe that the detected eye center always falls within the pupil; i.e., the solid black curve reaches 100% for \( nd \leq 1 \) (the exact value is 99.5%). Moreover, for most images, the detected point is relatively close to the pupil center (\( nd \leq 0.3/0.4 \) for ~80/90% of the database, respectively). On the other hand, with the database acquired under less-controlled conditions (MobBIO), the detected eye center falls within the pupil in about 77% of the images; it is worth noting, however, that the detected point is within the sclera for nearly the whole database (the exact figure is 97.4%). Some examples of eye detection in images from MobBIO are given in Figure 6. The three examples where detection occur in the vicinity of the eye center (top) show cases of off-angle image (left), occluded eye (center) and reflections due to glasses (right). The images below are examples of unsuccessful detection. Further examination of the first case (left) reveal that the inner (pupil) iris boundary is hardly visible and as a result, the filter response is weakened. The same happens in the second case (center) due to occlusion. The third example (right) shows a maxima in the region of interest, but a stronger maxima occur due to curve-shaped boundaries given by the glasses of the contributor.

Next, we report verification and identification results using the periocular system. Due to different image size, filter wavelengths span the range 4-16 with MobBIO and 16-60 with BioSec. For each database, this covers approximately the range of pupil radius of all its images, as given by the groundtruth. We evaluate two different grid configurations (Figure 3), one with dense image sampling and another with coarse sampling. The parameter \( d \) in Figure 3 indicates the distance between sampling points. The dense grid with BioSec has \( 13\times19=247 \) points, and the coarse grid has \( 7\times9=63 \) points. Similarly, the dense grid with MobBIO has \( 9\times13=117 \) points, and the coarse grid has \( 5\times7=35 \) points.

Results of the verification and identification experiments are given in Figure 7. Recognition experiments are reported in two scenarios: i) using as input the groundtruth pupil center, and ii) using the detected eye position given by our automatic detection system. Results of the iris matcher are reported using the noise mask given by the groundtruth, as in Figure 4. Finally, we also report results of the fusion between the iris texture matcher and the periocular matcher. For this purpose, we use the periocular configuration with coarse sampling and groundtruth information. It is observed in our experiments that rotation compensation does not have appreciable effects in the performance. This would allow to save computational time by suppressing this step. Given this result and for the sake of simplicity, from now on we will refer only to the case without rotation compensation.

It can be observed that with BioSec, results with automatic eye detection are quite similar to using groundtruth.
Fig. 6. Examples of eye detection. Top: successful localization in the vicinity of the eye center. Bottom: unsuccessful localization. Image $I_{20}$ is also given. The detected eye center is marked with a prominent red dot. Additional red points indicate other local maxima. The green dots are the pupil/sclera centers marked manually. All images are from MobBIO.

This is consistent with the higher accuracy shown by the eye detection system with this database. There are also no appreciable differences between the dense and coarse grid, which is a very good result considering that the dense grid has four times more points. The worse accuracy of the eye detection system with MobBIO is reflected in a reduction in performance when using automatic eye detection, as observed in the verification and identification curves. Interestingly enough, there is less reduction in accuracy with the coarse grid in verification mode (DET curves are closer) and in identification mode with a small hit list (Top1-5). An explanation can be that since the dense grid has more points which are closer to each other, it is more sensitive to spatial displacements given by a errors in the detection of the eye. Again, here there are advantages in working with a coarser grid.

As regards to the differences in performance between the two databases, performance in general is better with BioSec, specially in verification. This could be because images in MobBIO are more than half in size, resulting in a loss of identity information, and acquired under more adverse conditions. Other studies using bigger periocular images suggest, however, that visible images (as those in MobBIO) are better suited for periocular recognition that NIR images [4].

With respect to the iris matcher, its performance is much better in BioSec than in MobBIO. This is expected, since iris systems usually work better in NIR range [19]. An additional factor could be the differences in image size between the two databases, and the worse acquisition conditions of MobBIO. It is of relevance that the periocular system works better than the iris matcher in MobBIO. The small image size makes more difficult to reliably extract identity information for the iris texture. When this happens, the periocular region is still able to provide a rich source of identity data. Also, even in the adverse acquisition conditions of MobBIO, the iris system is able to complement the periocular system, as shown in the fusion results. On the other hand, the big difference in performance between iris and periocular found in BioSec can be one reason that explains why the fusion of the two systems does not improve performance (and it even worsens). The latter, however, should not be taken as a general statement. Other fusion rules may lead to different results with BioSec, specially if the supervisor is data quality and/or expert adaptive [28, 26]. Working towards the improvement of our periocular system can be another avenue to overcome this issue.

7. CONCLUSIONS

An eye detection system based on symmetry filters is presented. It is based on 2D separable symmetry filters tuned to detect circular symmetries, in such a way that we detect the eye center with a few 1D convolutions. This detection system is used as input to a periocular algorithm based on retinotopic sampling grids and Gabor analysis of the power spectrum. This framework is evaluated with two databases of iris data, one acquired with a close-up NIR camera, and another in visible light with a webcam. Despite the worse accuracy of the detection system with the webcam database, the periocular system is shown to be robust to inaccuracies in detecting the eye center, making the proposed setup suitable for unconstrained working conditions. In addition, the accuracy of the periocular system is not jeopardized by reducing the density of the sampling grid, allowing additional time savings. We also evaluate an iris texture matcher based on 1D Log-Gabor wavelets. Despite the poorer performance of the iris matcher with the webcam database, its fusion with the periocular system results in improved performance. Future work includes evaluating the reliability of the proposed eye detection system in full face images. Another source of improvement will be the incorporation of a refinement stage, e.g. by pixel-wise analyzing the neighborhood of the detected point to achieve a more accurate estimation of the eye center [2].

Acknowledgments

F. A.-F. thanks the Swedish Research Council and the EU for for funding his postdoctoral research. Authors acknowledge the CAISR program of the Swedish Knowledge Foundation, the EU BBfor2 project and the EU COST Action IC1106. Authors also
would like to thank the Biometric Recognition Group (ATVS-UAM) for making the iris part of the BioSec database available for our experiments.

8. REFERENCES


