Gaze-angle Impact on Iris Segmentation using CNNs

Ehsaneddin Jalilian and Andreas Uhl
University of Salzburg
Jakob-Haringer-Str.2, Salzburg, Austria
ejalilian@cs.sbg.ac.at, uhl@cs.sbg.ac.at

Mahmut Karakaya
University of Central Arkansas
201 Donaghey Ave, Conway, Arkansas, USA
mkarakaya@uca.edu

Abstract

*Emerging standoff iris recognition systems operate under unconstrained conditions and the iris images captured by these systems are more subject to off-angle acquisition distortions. While deep learning techniques (*e.g. *convolutional neural networks (CNNs)) are increasingly becoming a tool of choice for iris segmentation tasks, yet there is a significant lack of information about how these distortions affect the performance of such networks. In this work, we thoroughly discuss the general effect of different gaze-angles on ocular biometrics and relate the findings to off-angle iris segmentation using CNNs. In particular, we conduct systematical analysis on the impact of different gaze-angles on segmentation performance of two CNNs with different architectures. The networks' performance turns out to have a direct relation to the closeness of gaze-angles in the training and testing images, and it declines as the gaze-angles diverge. We further investigate the effect of (i) increasing the quantity of iris training data in case of gaze-angles in training and test data match, and (ii) considering iris training data consisting of several distinct gaze-angles (we obtain promising results using the second configuration). Finally, we compare our results to those of some classical iris segmentation algorithms, where the CNNs are found to outperform the classical algorithms.*

1. Introduction

Iris recognition is a method of identifying people based on unique patterns within the ring-shaped region surrounding the pupil of the eye. Like any other biometrics system, the performance of iris recognition systems is highly dependent on accurate segmentation of the target region (iris texture) from the rest of the image. Existing iris recognition systems are designed to capture the iris image mostly when the iris plane is almost perpendicular to the visual axis of camera. In this case, as sample images from which templates are generated are captured almost under the same conditions and have similar distortion effects such as occlusion, focus, motion-blur, and specular reflections, the segmented iris region in the probe image corresponds to that of the gallery image(s), maximizing the correlation between the extracted features. Recent demands for standoff iris biometric systems and the trend towards "on-the-move-acquisition" are transforming iris biometric systems from being operated in a well-established and well-controlled setup, to being a smart standoff modality. Due to their much less constrained setup, iris images captured by these systems are more likely to be off-angle, and thus incorporate additional off-angle related distortions. Therefore, off-angle iris recognition has became an emerging research topic in biometrics recently. Showing superior performance over classical methods, convolutional neural networks (CNNs) represent a new paradigm gaining increasing interest in the biometrics community. To this extent, recently some CNN models got proposed for iris segmentation as well [9]. While there are already several approaches addressing the off-angle distortions in classical iris segmentation methods, yet there exists no detailed research that investigates and quantifies the mechanism and extent of off-angle, eye-structure related distortions on the segmentation capability of the CNNs.

In this work, we cover this topic by investigating the effect of different gaze-angles in ocular biometrics in general and relating the gained information to off-angle iris segmentation using CNNs. This allows us to figure out the proper segmentation strategy when dealing with iris images captured from different angles (*e.g.* captured by standoff systems). Eventually, it might turn out that CNN training has to be done specifically for a certain gaze angle (i.e. employing annotated iris images of this angle for training only - *homogeneous* training data). In that case, in the iris recognition application context, we may utilize available eye-gaze estimation algorithms [14, 17] to classify acquired iris images based on their gaze-angle, and next conduct segmentation using the appropriately trained CNNs (training each CNN with data of one single gaze-angle only). This approach requires the availability of a significant number of appropriately trained CNNs. This procedure is not desirable. Al-

978-1-7281-1522-1/19/$31.00 ©2019 IEEE

ternatively, it would be of advantage to improve the generalization capability of the networks, and thus eliminate the need for any further processing stages including gaze-angle estimation and angle-specific training.

To address these research questions, we first train two semantic segmentation oriented fully convolutional neural networks (FCNs) with different architectures, using iris images acquired from eleven different angles in a homogeneous training scenario. We then analyze their segmentation performance on iris images captured from angles matching as well as not-matching the training data. We further investigate two training settings which are based on (i) heterogeneous-angle training (training set contains iris images with different gaze angles), and (ii) increasing the size of the homogeneous training sets, to improve the generalization capability of the networks. Finally, we evaluate the networks' segmentation accuracy as compared to some classical iris segmentation algorithms applied on the same data. To the best of our knowledge, this is the first work which conducts such investigations systematically in ocular biometrics in the context of iris segmentation using CNNs.

2. Related work

Existing iris segmentation methods can be broadly divided into four types. The first and most popular type is feature-based methods. These algorithms first find the pupil and the inner boundary of the iris, and after locating the outer iris boundary, articulate the other parameters, such as eyelid and limbic areas, to separate them from the iris. Two well know algorithms in this category are Hough transform (HT) and Daugmans integro-differential operator. HT finds the circularity by edge-map voting within the given range of the radius, which is known as the Wildes approach [30]. Daugmans integro-differential operator is another scheme that finds the boundary using an integral derivative, based on which new advanced methods got developed recently [1]. An effective technique to reduce the error rate in a non-cooperative (*e.g.* off-angle) environment was proposed by Jeong et al. [27]. They used two circular edge detectors in combination with AdaBoost for pupil and iris boundary detection, and their method approximated the real boundary of detected eyelashes and eyelid. Other methods are also known to reduce noise prior to the detection of the iris boundary to increase segmentation accuracy [28]. Performance of these methods is highly dependent on the images' clear contour and the boundaries' contrast as well as the circular/elliptic shape of the pupil and iris, however in off-angle images limbic and pupillary boundaries are usually of uneven-contrast, and have non-circular/elliptic shape.

The second type of methods includes texture-based segmentation schemes. These algorithms use the specific color texture and the illumination information gradient to differentiate between an iris pixel and another pixel. Thus, iris segmentation is performed based on the discriminating features for iris and non-iris pixels. A novel method for iris and pupil segmentation using this technique was proposed by Khan et al. [16]. They used 2-D profile lines between the iris and sclera boundary and calculated the gradient pixel by pixel, where the maximum change represents the iris boundary. Parikh et al. [22] first approximated the iris boundary by color-based clustering, then for off-angle eye images, two circular boundaries of the iris were detected.

The third type of segmentation methods employs active contour methods [26]. In the VanChese algorithm a mask is created according to the size of the iris, and then an iterative process determines the true iris boundary with the help of the localized region-based formulation [31]. However, this approach shares the drawback faced by other active contour-based models, as it is usually disturbed by the iris texture during iteration, and normally considers the iris pattern as the boundary, which results in inaccurate segmentation. Due to the space limitation here we just introduced a selection of techniques in each category, however there exists current and ongoing research on non-ideal and off-angle iris segmentation using the classical approaches specified. For a general overview please refer to *e.g.* [13] and [5].

Addressing the drawbacks of classical segmentation methods and reducing the complexity of intensive pre- and post-processing, a fourth category of segmentation methods evolved recently, which are based on data-driven learning methods. Within this category, deep learning techniques and in particular convolutional neural networks are the most ideal and popular schemes due to their accuracy and performance. Liu et al. [20] located the iris region in non-cooperative environments using convolutional neural networks. In their study, a hierarchical CNN (HCNNs) and a multi-scale FCN (MFCNs) were used to locate the iris region automatically. Jalilian and Uhl [9] proposed three types of fully convolutional encoder-decoder networks for iris segmentation, and evaluated their performance on off-angle iris images available in UBIRIS.v2 database[1]. Their results showed the superior capability of CNNs to deal with off-angle iris data compared to some classical methods.

Arsalan et al. [19] proposed a two-stage iris segmentation based on CNNs for images captured in visible light. Authors used circular Hough transform to detect the rough iris boundary in the first stage. A pre-trained VGGface model is used in the second stage for the fine adjustment of the rough iris boundary obtained in the first stage. Osorio-Roig et al. [21] conducted inductive learning over two FCNs for segmenting several eye regions in multi-class approaches, using fine-tuned pretrained VGG-16 and AlexNet models. Similarly, Rot et al. [25] presented a deep multi-class eye segmentation model built around a semantic segmentation architecture. They have also examined sensitiv-

[1]http://iris.di.ubi.pt/ubiris1.html

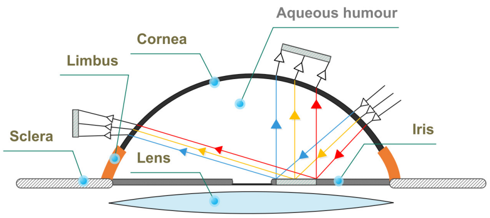


Figure 1: Posterior eye structure and the perspective and refraction distortions affecting the iris texture geometry.

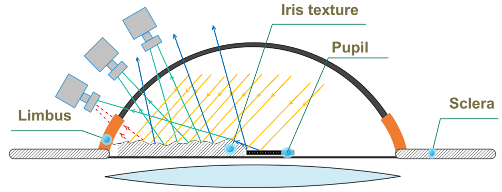


Figure 2: Three-dimensional structure of iris and Limbus occlusion distortions

ity of the network to the change of view for four directions (left, right, up and straight) generally. In order to overcome the requirement of large quantities of labeled data in the approaches mentioned above, Jalilian et al. [12] proposed a domain adaption technique for CNN based iris segmentation. Bazrafkan et al. [4] introduced a CNN to perform iris segmentation on lower-quality iris images (including off-angle images). They further investigated the effect of network tuning on the segmentation results. Nevertheless, none of the above works provided a systematic analysis on the effect of the different gaze-angle on ocular biometrics and the resulting iris segmentations using CNNs.

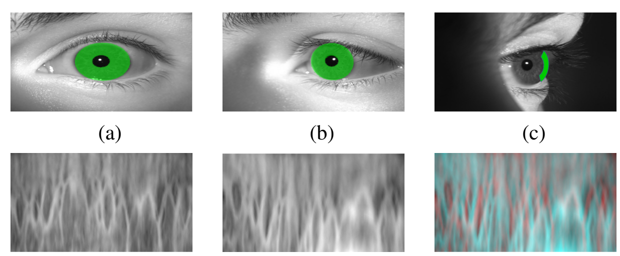
3. Eye structures effect on iris segmentation

In addition to the known degradation factors affecting the constrained (frontal) iris imaging (such as pupil dilation, occlusion, image resolution, focus, motion blur, specular reflections, and illumination variations), off-angle iris imaging introduces further challenging eye-structure related distortions to the iris images, including perspective and refraction distortions, change in the appearance of complex three-dimensional texture on the iris plane, iris missing boundary in extreme angles, and limbus occlusion.

Interaction of light rays within the posterior eye structure elements such as cornea, limbus, sclera, anterior chamber (aqueous humour), iris, and lens (as illustrated in Figure 1) can distort the actual iris image depending on the image acquisition angle. In both frontal and off-angle images, the cornea and aqueous humour first refract the incoming and outgoing light rays based on their angles to the cornea. Further perspective distortions may get introduced to the light rays as the perspective (gaze-angle) changes. The three-dimensional texture on the iris plane can appear differently as the angle changes and creates shadows on the iris plane. Also in extreme angles the sclera-iris boundary disappears, distorting the iris' actual circle boundary shape. After all, the limbus, which is a semi-transparent structure at the junction of the cornea and sclera, consistently occludes side portions of the iris plane. In any case, the extent of distortion in the iris texture depends on the gaze-angle of the iris images.

3.1. Three-dimensional structure of iris

The structure of the iris consists of several types of dilator muscles to contract the pupil to control the amount of the incoming light to the eye lens, forming a three-dimensional texture on the iris plane. To this extent, the key context and texture feature representations learned by the CNN networks change from a certain view angle to another. For instance, some iris pixels may get occluded or shadowed by others and consequently, the 2D image of the captured iris texture changes as the gaze-angle changes (see Figure 3f, which shows the difference between the normalized images of an iris captured frontally (Figure 3d), and from +50° gaze-angle (Figure 3e) in red, where constant parts are depicted in several shades of blue). Those pixels located on (the side closer to the camera) the border region of the iris inner boundary may get occluded in steeper view angles (see the light rays unseen (blue) and seen (green) by the camera in different angles in Figure 2). There will be considerable changes in the distribution of iris features when the gaze-angle of training and testing iris images differs, and the network may not be able to spot the corresponding features (as learned in the training session) in the test images, failing to segment the iris region accurately.



(a) (b) (c)

(d) (e) (f)

Figure 3: Examples of off-angle distortions on the iris images in the database.

3.2. Limbus occlusion

The limbus is the semitransparent organ that joins the sclera and the cornea texture where the fully transparent cornea cannot reach to the bottom of the anterior chamber and ends at a higher level than the iris plane. Due to the distance between the ending points of the cornea and the iris plane, the diameter of the cornea-limbus border is slightly smaller than the anterior chamber width. Therefore, the limbus consistently occludes the boundary region of the iris texture (especially) in extreme off-angle view. The extent of occlusion of the iris texture on the side closer to the camera increases as the gaze angle increases (see the red-dotted reflections in Figure 2). In CNN-based segmentation, an off-angle test image does not exhibit certain outer iris boundary information as present in frontal training images, and thus the network fails to accurately detect the iris region (especially its outer boundary) in the test images.

3.3. Perspective and refraction distortion

The geometric properties of an object's image on the camera sensor change if the coordinates of the camera change with respect to the object. This phenomenon is simply referred to as "perspective distortion". In this case the new position of point a_{x_1,y_1,z_1} rotated by $\theta_{x,y,z}$ with respect to a coordinate system defined by the camera will be located at ${a'}_{x_2,y_2,z_2}$ as follows:

$$\begin{aligned} {a'}_{x_2} &= c_y(s_z y_1 + c_z x_1) - s_y z_1 \\ {a'}_{y_2} &= s_x(c_y z_1 + s_y(s_z y_1 + c_z x_1)) + c_x(c_z y_1 - s_z x_1) \\ {a'}_{z_2} &= c_x(c_y z_1 + s_y(s_z y_1 + c_z x_1)) - s_x(c_z y_1 - s_z x_1) \end{aligned} \quad (1)$$

where c and s stand for cos and sin of the rotation degree (θ), respectively. Figure 1 illustrates the effect of this phenomenon on iris images as the capturing angle changes.

The cornea is the transparent structure of the eye located at the outermost layer of the eye. Aqueous humour is the transparent watery fluid that is located between the cornea and the iris and fills the anterior chamber. Therefore, incoming and outgoing light rays are first refracted at the cornea and then refracted at the aqueous humour due to the refraction index differences between air, cornea, and aqueous humour. When capturing iris images at steeper angles, light rays refract more at the cornea, causing the geometric property of the reflected iris features to transform (*e.g.* get scaled, dilated or eroded), as shown in Figure 1. Correspondingly, we can see that the circle shape of the iris image captured in frontal manner in Figure 3a is transformed to an ellipse in Figure 3b, when captured from a +50°angle, mainly due to the perspective distortion. We can also observe the effect of this distortion along with the refraction distortion on the geometric properties of the corresponding normalized iris textures (Figure 3d, and Figure 3e respectively), as presented in Figure 3f. Basically CNNs learn scale-dependent patterns at a specific combination of image size and network architecture, and thus they are not able to spot the learned patterns in the testing data, if their geometric properties (such as boundaries and texture information) are changed with respect to the training data.

3.4. Iris missing boundary in extreme angles

The sclera is the outer layer of the eye with bright white color which strongly contrasts with the colored iris texture, forming a clear boundary between these two tissues. In frontal imaging this boundary is clearly visible. But

Network	RefineNet	SegNet
Optimizer	Adam	Stochastic gradient descent
Learning rate	0.0001	0.003
Momentum	-	0.01
Weight decay	0.1	0.000001
Epochs	40,000	30,000

Table 1: Networks' training parameters

as the gaze-angle gets steeper (especially towards the right most gaze-angle as we consider left eyes (*i.e.* +50°)), the boundary erodes and finally disappears (see the green curve, showing the missing iris boundary in Figure 3c). The learning process in CNNs starts with convolving filters which can be thought of as feature identifiers which convolve over the input looking first for low level features such as edges and boundaries, and then building up to more abstract concepts through further filtering layers. Thus, low level features such as edges (boundaries) play a scaffolding role in encoding the feature representations of target regions. Therefore, if these features (learned during the training process) are not presented in the testing data (which might have steeper gaze-angle than the training data), the network will not be able to retrieve the accurate boundary pixels (which are missing), or may spot false boundaries (false-positives) in the image wrongly.

4. Experimental framework

Database: For our experiments we used a subset (containing 4400 left eye iris images captured from 40 subjects) of an off-angle iris database [15]. The iris images in this database are captured by two near-infrared sensitive IDS-UI-3240ML-NIR cameras. Images at 0°gaze-angle were captured by a frontal fixed camera, and off-angle images were captured by a frontal moving camera rotating horizontally from -50°(N50) to +50°(P50) in angle with a 10°step-size. Each camera captured 10 iris images per stop, giving 10 frontal and 100 off-angle iris images captured from each subject, to comprise 440 images per angle (examples of images in the database are presented in Figure 6). The database is accessible on request (from the authors), and further details about it can be found in [15]. We developed the ground-truth segmentation masks (required for training networks) for all images available in the dataset using the iris, pupil, upper and lower eyelid parameters specified manually. For our experiments we divided the whole database into two equal parts (each containing iris images of 20 separate subjects), and used one part as our testing data and the other one as our training data. Jalilian et al. already showed in their works that the CNN networks can be trained to their optimal accuracy using approximately 200 training samples [10] [11].

Segmentation evaluation and measures: In order to facilitate proper quantification of the resulting segmentation accuracy in each experiment, we considered the $nice1$ iris

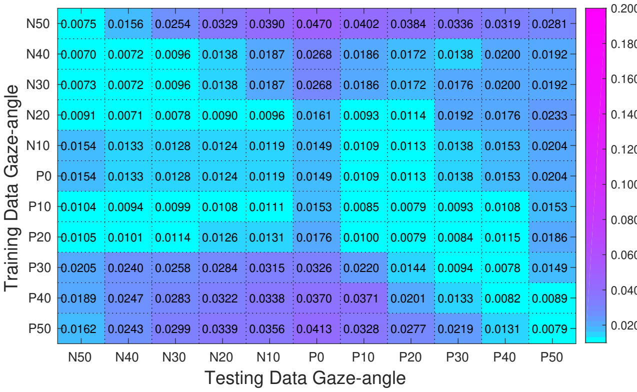


Figure 4: Segmentation performance of SegNet trained and tested on each gaze-angle separately as average $nice1$ error.

segmentation error rate which is based on the NICE1 protocol [2] as used in several iris segmentation challenges. Accordingly, the segmentation error rate (E_i) for each input iris mask I_i is given by the proportion of corresponding disagreeing pixels (through the logical exclusive-or operator) over all the mask as follows:

$$E_i = \frac{1}{c \times r} \sum_{c'} \sum_{r'} O(c^{'}, r^{'}) \otimes C(c^{'}, r^{'}) \qquad (2)$$

where c and r are the dimensions of the segmentation masks, and $O(c', r')$ and $C(c', r')$ are, respectively, pixels of the output (segmentation result) and the ground-truth masks. The value of (E) is in the [0, 1] interval, and 1 and 0 are the worst and the best scores, respectively.

Fully convolutional neural networks (FCNs): In this work we used two different FCN architectures to extract the iris textures from iris images. The first network architecture we used is identical to the "Basic" fully convolutional encoder-decoder network proposed by Badrinarayanan et al. [2] and is termed "SegNet" subsequently. The network has a rather shallow structure, which includes an encoder net, and a corresponding decoder net. The network's encoder architecture is organized in four stocks, containing a set of blocks. Each block comprises a convolutional layer, a batch normalization layer, a ReLU layer, and a Pooling layer with kernel size of 2×2 and stride 2. The corresponding decoder architecture, likewise, is organized in four stocks of blocks, whose layers are similar to those of the encoder blocks, except that here each block includes an up-sampling layer. The decoder net ends up to a soft-max layer which generates the final segmentation mask. The network implementation[3] was realized in the Caffe framework.

The second network architecture we used in our work is RefineNet [18]. RefineNet has a very deep multi-path refinement architecture, which employs a 4-cascaded architecture with 4 Refining nets, each of which directly connects to the output of one Residual block [6], as well as to the preceding Refining net in the cascade. Each Refining net consists of two residual convolution units (RCU), and their outputs are fused into a high-resolution feature map, and then fed into a chained residual Pooling block. The implementation[4] of this network was realized in the TensorFlow framework using the Keras library. We selected these two architectures based on the successful results already obtained by these two networks in iris segmentation [7, 8], and also due to their architectural distinctions. None of the networks war pre-trained. Table 1 specifies the training parameters we used for the networks in our experiments.

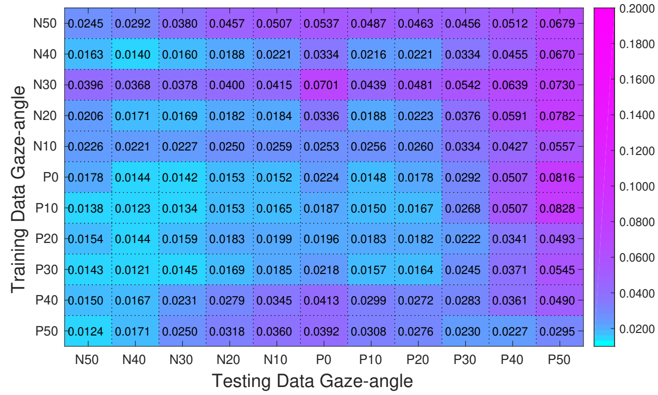


Figure 5: Segmentation performance of RefineNet trained and tested on each gaze-angle separately as average $nice1$ error.

5. Experiments and Results

The first research question addressed is if a gaze-angle specific training process is required for high segmentation accuracy. To facilitate a proper analysis, we trained the networks in a homogeneous training setting, i.e. training data consists of iris images with identical gaze angles (200 images per gaze-angle as in our training data). For each available gaze-angle, a dedicated network is trained. Subsequently, we conduct segmentation on all the test data, differentiating and grouping results into the different gaze-angles available (starting from -50°(N50) to +50°(P50)). Figure 4 shows the results (as average $nice1$ error) for this experiment for the SegNet network. As it can be seen, the network's performance has a direct relation to the similarity of gaze-angles of the training and testing images, and as the gaze-angles diverge, the performance of the network declines too. This result is underlined by the coloring, which displays the lowest errors around the diagonal, which reflects the effect of off-angle distortions explained in Section 3. As the corresponding outputs in Figures 6l and 6m also show, the missing left iris boundary and dilated right boundary (caused by perspective and refraction distortions) in extreme P50 gaze-angle images cause the network to fail

[2] http://nice1.di.ubi.pt/

[3] http://mi.eng.cam.ac.uk/projects/segnet/tutorial.html.

[4] https://github.com/eragonruan/refinenet-image-segmentation.

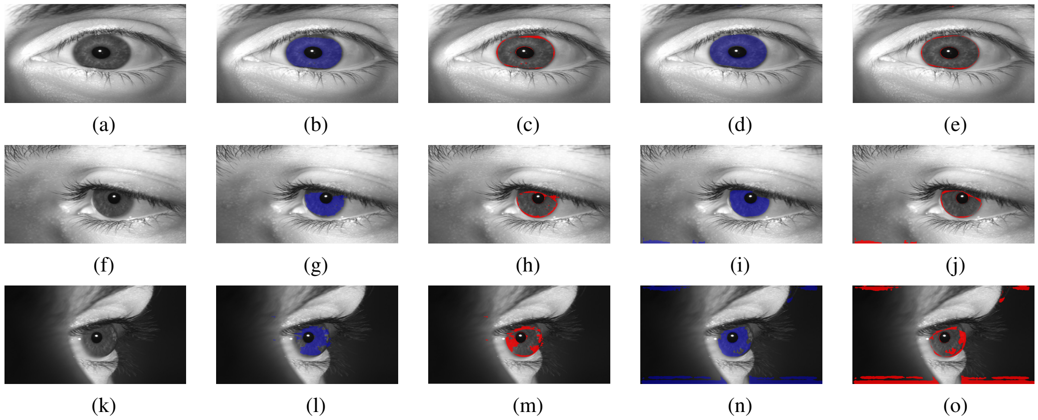


Figure 6: Sample iris images of P0 (6a), N50 (6f), and P50 (6k) gaze-angles in the database, and their corresponding segmentation (blue color) masks, and error (red color) masks using SeNet (second and third column), and RefineNet (fourth and fifth column), trained on frontal iris data, respectively.

to extract iris boundary pixels properly. Similar poor results are visible in the outputs of the N50 gaze-angle images (see Figures 6g and 6h), but with much less severity. The reason for this are the more severe off-angle distortions present in the images of P50 gaze-angle (compare the corresponding images in Figures 6f and 6k) which is not surprising as we consider left eyes only. As it is visible in the corresponding example outputs as well, the right iris boundary is also affected by the limbus occlusion which resulted in additional false-negative detections in this area. The other false-negative detections visible in the iris main texture (far from the iris boundaries) seem to be mainly due to the 3D iris structure, perspective and refraction distortions, affecting the iris texture property as the gaze-angle gets steeper (*i.e.* in P50 angle). As results of these distortions, we can observe relevant degradations in the segmentation results obtained by the network (as presented in Figure 4), which are visible in the upper right and the lower left corners of the table (corresponding to the configurations where the difference between the training and testing gaze-angles is larger).

Figure 5 shows the results of the same experiments using the RefineNet network, which look partially different than the results obtained by the SegNet network. Overall we observe higher errors, the better results when being close to the diagonal are not seen that clearly, and the table is much less symmetric. As already mentioned, RefineNet uses a chain of pooling filters with plenty of residual information in a very deep architectural configuration to build many more abstract concepts (contents) on the low level features (texture), aiming to enable high-resolution boundary predictions. To this extent, key content information (boundary information) plays a key role in adjusting the network's filter weights, and thus learning the target features. The corresponding output in Figure 6o also shows, that this architecture makes the network very vulnerable to the texture related distortions such as perspective, refraction and also 3D iris structure distortions. Correspondingly, we can observe many false-positive detections in the output mask, as well as some undetected iris pixels (false-negatives), especially on the iris main texture, when testing the network on iris images acquired from steeper angles (*e.g.* P30, P40 and P50). The strong decline of the network performance visible at the upper left and right side of the table is mainly due to this issue. These distortions affect the iris images of the left sided gaze-angles less severely. As already specified, the reason for this effect is the more severe off-angle distortions present in images of P50 gaze-angle (see Figures 6g and 6h), and thus we can observe better performance in the lower left and left side of the table. The effect of missing iris boundary is also visible in Figure 6o, however, the effect of limbus occlusion distortion seems not to so sever.

Network-wise, SegNet has a comparably shallow hierarchical encoding-decoding architecture which allows the network to extract a balance combination of both high level (content) and low level (texture) iris features up to a moderate level. While such a configuration allows the network to resist against false-positive errors (caused by the off-angle distortions) when trained on frontal images and tested on off-angle images, it undermines the network capability in detecting the iris pixels located at the boundary regions properly (resulting in many false-negative detections). On the contrary, RefineNet has a very deep architecture with built-in chain of pooling filters to extract more abstract concepts in the higher layers. Thus, the network is very sensitive to the higher level feature information. While the network performance is good in configurations where the effect of off-angle distortions on the texture pixels are not very severe, it diminishes strongly (spotting many false-negative pixels) when such distortions are high. Consequently, SegNet shows better accuracy in such configurations.

The next research question tackles the impact of the size of the training set and the stability of the obtained results

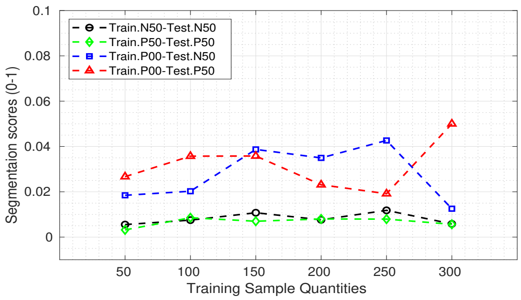


Figure 7: Segmentation performance of SegNet (average $nice1$), trained with increasing quantities of P0 gaze-angle images, on the N50 and P50 gaze-angles images *vs*. trained with identical (N50 and P50) gaze-angles images.

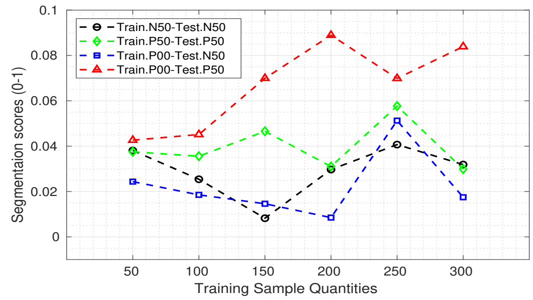


Figure 8: Segmentation performance of RefineNet (average $nice1$), trained with increasing quantities of P0 gaze-angle images, on the N50 and P50 gaze-angles images *vs*. trained with identical (N50 and P50) gaze-angles images.

with respect to a varying quantity of training images. In particular we wanted to clarify if the results get better (i.e. improved generalizability in case training and testing angle do not match) by simply increasing the number of used training images. For this purpose, we trained the networks with increasing quantities (50, 100, 150, 200, 250, and 300) of N50 and P50 gaze-angle images, respectively, and then tested them on the remaining (100) images of the same gaze-angles. Similarly, we trained networks with increasing quantities of frontal images (P00), and then tested on the same N50 and P50 gaze-angle images as before. Figures 7 and 8 demonstrate the results obtained using SegNet and RefineNet for these experiments, respectively. As the results show, in general introducing more training images to the networks doesn't consistently improve the networks segmentation performance. Moreover, results indicate very unstable results and only for SegNet we observe stable accuracy for varying training data size in case training and testing gaze angle do match. In all other configurations (especially those involving RefineNet) we observe highly unstable behavior, often showing increasing error for larger training sets. Therefore, the results obtained so far do suggest that (i) in fact gaze angle-specific training is required to get high accuracy and that (ii) increasing the size of the training set does not resolve this issue. In particular for RefineNet, we observe very unstable results when changing the size of the training set.

The next research question we want to address is if we are able to improve the generalizability of the networks by switching to a heterogeneous training setting, i.e. to include iris images with different gaze-angles into the training set. For this purpose, in the first stage, we trained the networks with all available iris images (with different gaze-angles) in our training data (200 samples per gaze-angle), and then tested the networks on all images in the testing data. Figure 9 visualizes results for this experiment per gaze-angle. As it can be seen, the networks react quite differently to this training configuration. SegNet's performance gets almost identical to the results of the homogeneous training setting in case training and testing gaze-angles do match. This is a very positive result, as this enables us to refrain from the angle-specific training strategy (and even better, from the application perspective, there is no need to determine the gaze-angle of an iris image before being able to deploy CNN-based segmentation). However, RefineNet's performance deteriorates considerably as compared to these best results but is still better as compared to homogeneous training with not corresponding gaze-angles (see Figures 4 and 5). Actually, RefineNet seems to be unable to properly learn the features when introducing the diversified target features as contained in the heterogeneous training set due to its concentrated learning mechanism already explained.

This first experiment with heterogeneous training sets was based on a large training set. In the next stage, we want to investigate how many training images per gaze-angle we actually need to achieve results of this high accuracy. For this, we trained the networks with different shares of available iris images and present results when using the half quantity (100 samples per gaze-angle in our training set) – this setting is the best for RefineNet while for SegNet the full training set is optimal. As expected and in accordance to earlier results, SegNet's accuracy gets eroded considerably, but surprisingly RefineNet's accuracy improves notably (getting even better than the results of the homogeneous training setting on the left-sided gaze-angle iris im-

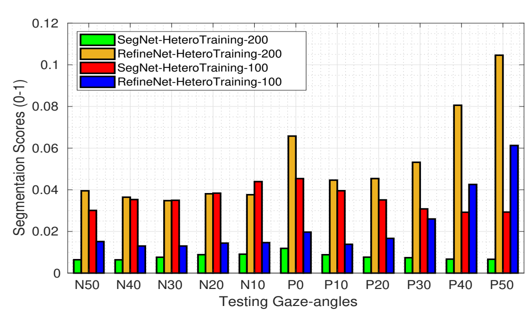


Figure 9: Segmentation performance of SegNet and RefineNet, trained with iris images of different gaze-angles altogether as average $nice1$ error.

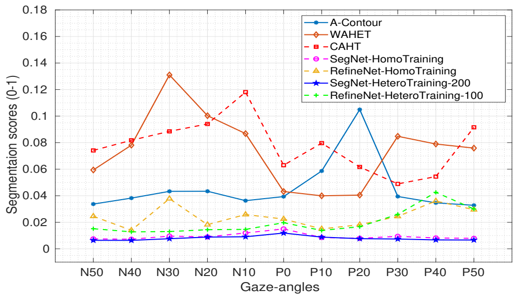


Figure 10: Segmentation performance of the CNN networks *vs*. classical methods' performance as average $nice1$ error.

ages). The network's concentration on high level content information can explain this phenomenon. Actually, the training images are (supposed to be) originating from the same feature space (sharing similar contents). Thus, the discrepancies among differing angles are mostly embedded in the texture data. Therefore, introducing more training samples to the network means introducing more discrepancies to the network (which is not designed to handle them well, resulting in worse accuracy for the higher quantity of heterogeneous training data). Overall, we can conclude that training a (certain) network architecture with a distinct quantity of iris images of different gaze-angles improves the network's generalization capability and its subsequent segmentation accuracy. However, if we select an improper architecture or quantity of training data, the results may not be as expected.

Finally, the last research question is to evaluate the actual segmentation performance of the networks on iris images with different gaze-angles as compared to classical segmentation techniques. To address this, we compare the corresponding segmentation results obtained by the networks on each angle to those obtained by applying three well known classical iris segmentation algorithms, including: active contours-GrabCut (A-Contour) [3], contrast-adjusted Hough transform (Caht) [24], and weighted adaptive hough and ellipsopolar transform (Wahet) [29]). We used the USIT implementation of these algorithms [23]. To enable a fair comparison, we used the results obtained by training the networks with heterogeneous gaze-angles in their best configuration (i.e. large training set for SegNet and small training set for RefineNet) (Figure 9), as well as those obtained by using the homogeneous training setting with full training set for both networks (see Figures 4 and 5). As the results in Figure 10 show, both networks have superior performance over the classical methods, no matter if trained on the homogeneous or heterogeneous data (best configuration), while SegNet shows consistently better performance.

6. Conclusion

We investigated the effect of different gaze-angles on ocular biometrics and related the obtained information to off-angle iris segmentation using CNNs. The results showed that the performance of the networks has a direct relation to the correspondence of the gaze-angles of the training and testing images, and it declines as the gaze-angles diverge, confirming the negative effect of off-angle eye-structure related distortions on the networks' performance. While missing and dilated iris boundary distortions (especially in P50 gaze-angle images) caused the networks to fail to accurately extract the iris boundary pixels, the effect of perspective, refraction, and 3D iris structure distortions on the networks mainly appeared as missing (undetected) iris texture in the networks' output masks. Also limbus occlusion affected the networks' performance, mostly resulting in missing iris outer boundary pixels (false-negatives) in the corresponding segmentation masks of images with steeper gaze-angles. In this case, the effect of distortions was more severe on the images captured from the right angles, rather than the images captured from the left angle, due to the more severe off-angle distortions they include when captured from the right angles, as we consider left eyes only. It also turned out that FCNs with shallow hierarchical architecture, which extract a balanced combination of both high and low level features up to a moderate level, resist more against off-angle distortions (when gaze-angles in training and test data do not match), generating fewer false-positive errors. In contrast, FCNs with very deep architectures (designed to extract more abstract features) are very sensitive to the off-angle distortions affecting the texture data, thus detecting many false-positive pixels.

The investigations on the stability of the results and the possibility of improving them by increasing the quantity of training images turned out to not deliver promising outcomes for none of the networks, as there was no converging pattern observed when increasing the training images' quantity. Alternatively, we considered to train the networks with iris images exhibiting different (heterogeneous) gaze-angles, aiming to increase the networks' generalization capability in such cases. The experimental results showed that this configuration can enhance a network's generalization capability and significantly improves the results, especially for SegNet. Gaze-angle specific network training can be abandoned when resorting to this training strategy. However, we found that selecting a proper network architecture and training data quantity plays a key role in such a configuration. Last but not least, comparison of the corresponding segmentation results obtained by FCNs to those of some classical iris segmentation algorithms showed that FCNs possess superior performance over the classical algorithms, making them a better choice for off-angle iris segmentation.

Acknowledgement

This project received funding from the European Union's Horizon 2020 program under grant agreements No 700259 (PROTECT), 690907 (IDENTITY) and NSF grant No CNS-1909276.

References

[1] D. Ann-Roy and U. S-Soni. Iris segmentation using daughman's method. In *International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, pages 2668–2676, 2016.

[2] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.

[3] S. Banerjee and D. Mery. Iris segmentation using geodesic active contours and grabcut. In *Revised Selected Papers of the PSIVT 2015 Workshops on Image and Video Technology*, volume 9555.

[4] S. Bazrafkan, S. Thavalengal, and P. Corcoran. An end to end deep neural network for iris segmentation in unconstrained scenarios. *Neural Networks*, 106:79–95, 2018.

[5] M. De-Marsico, M. Nappi, F. Narducci, and H. Proença. Insights into the results of miche i-mobile iris challenge evaluation. *Pattern Recognition*, 74:286–304, 2018.

[6] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[7] H. Hofbauer, E. Jalilian, A. Sequeira, J. Ferryman, and A. Uhl. Mobile nir iris recognition: Identifying problems and solutions. In *International Conference on Biometrics: Theory, Applications, and Systems (BTAS2018)*, page 9, 2018.

[8] H. Hofbauer, E. Jalilian, and A. Uhl. Exploiting superior cnn-based iris segmentation for better recognition accuracy. *Pattern Recognition Letters*, 2018.

[9] E. Jalilian and A. Uhl. Iris segmentation using fully convolutional encoder–decoder networks. In *Deep Learning for Biometrics*, pages 133–155. 2017.

[10] E. Jalilian and A. Uhl. Finger-vein recognition using deep fully convolutional neural semantic segmentation networks: The impact of training data. In *International Workshop on Information Forensics and Security (WIFS)*, pages 1–8, Hong Kong, 2018.

[11] E. Jalilian and A. Uhl. Enhanced segmentation-cnn based finger-vein recognition by joint training with automatically generated and manual labels. In *International Conference on Identity, Security and Behavior Analysis*, pages 1–8, 2019.

[12] E. Jalilian, A. Uhl, and R. Kwitt. Domain adaptation for cnn based iris segmentation. *International Conference of the Biometrics Special Interest Group*, pages 1–6, 2017.

[13] F. Jan. Segmentation and localization schemes for non-ideal iris biometric systems. *Signal Process.*, 133(C):192–212, Apr. 2017.

[14] A. Kar and P. Corcoran. A review and analysis of eye-gaze estimation systems, algorithms and performance evaluation methods in consumer platforms. *IEEE Access*, 5:16495–16519, 2017.

[15] M. Karakaya, D. Barstow, H. Santos-Villalobos, and J. Thompson. Limbus impact on off-angle iris degradation. In *International Conference on Biometrics (ICB)*, pages 1–6, 2013.

[16] T. Khan, A. Khan, S. Malik, S. Khan, T. Bashir, and A. Dar. Automatic localization of pupil using eccentricity and iris using gradient based method. *Optics and Lasers in Engineering*, 49(2):177–187, 2011.

[17] J. Lemley, A. Kar, A. Drimbarean, and P. Corcoran. Efficient CNN implementation for eye-gaze estimation on low-power/low-quality consumer imaging systems. *CoRR*, abs/1806.10890, 2018.

[18] G. Lin, M. Anton, S. Chunhua, and I. Reid. Refinenet: Multipath refinement networks for high-resolution semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5168–5177, 2017.

[19] A. Muhammad, H. Hyung-Gil, N. R. Ali, L. Min-Beom, K. Min-Cheol, K. Dong-Seop, K. Chan-Sik, and P. Kang-Ryoung. Deep learning-based iris segmentation for iris recognition in visible light environment. *Symmetry*, 9(11):263, 2017.

[20] L. Nianfeng, L. Haiqing, Z. Man, L. Jing, S. Zhenan, and T. Tieniu. Accurate iris segmentation in non-cooperative environments using fully convolutional networks. In *International Conference on Biometrics (ICB)*, pages 1–8, 2016.

[21] D. Osorio-Roig, C. Rathgeb, M. Gomez-Barrero, A. Morales-González, E. G. Llano, and C. Busch. Visible wavelength iris segmentation: A multi-class approach using fully convolutional neuronal networks. In *International Conference of the Biometrics Special Interest Group*, pages 1–5, 2018.

[22] Y. Parikh, U. Chaskar, and H. Khakole. Effective approach for iris localization in nonideal imaging conditions. *IEEE Students' Technology Symposium*, pages 239–246, 2014.

[23] C. Rathgeb, F. Struck, and C. Busch. Efficient bsif-based near-infrared iris recognition. In *International Conference on Image Processing Theory, Tools and Applications (IPTA'16)*, 2016.

[24] C. Rathgeb, A. Uhl, and P. Wild. *Iris Recognition: From Segmentation to Template Security*, volume 59 of *Advances in Information Security*. Springer Verlag, 2013.

[25] P. Rot, Ž. Emeršič, V. Struc, and P. Peer. Deep multi-class eye segmentation for ocular biometrics. In *International Work Conference on Bioinspired Intelligence*, pages 1–8, 2018.

[26] S. Shah and A. Ross. Iris segmentation using geodesic active contours. *Trans. Info. For. Sec.*, 4(4):824–836, 2009.

[27] D. Sik-Jeong, J. Won-Hwang, B. Jun-Kang, K. Ryoung-Park, C. Sun-Won, D. Kwon-Park, and J. Kim. A new iris segmentation method for non-ideal iris images. *Image and Vision Computing*, 28(2):254–260, 2010.

[28] T. Tan, Z. He, and Z. Sun. Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition. *Image and Vision Computing*, 28(2):223–230, 2010.

[29] A. Uhl and P. Wild. Weighted adaptive hough and ellipsopolar transforms for real-time iris segmentation. In *International Conference on Biometrics (ICB)*, pages 1–8, 2012.

[30] R. Wildes. Iris recognition: an emerging biometric technology. *Proceedings of the IEEE*, 85(9):1348–1363, 1997.

[31] T. Yuen-Chai, B. Min-Goi, Y. Haur-Tay, W. Kiat-Chin, Y. Lung-Lai, and L. Kong-Chian. Local chan-vese segmentation for non-ideal visible wavelength iris images. In *Conference on Technologies and Applications of Artificial Intelligence (TAAI)*, pages 506–511, 2015.