

© IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE. This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder.

EXPLOITING SUPERIOR CNN-BASED IRIS SEGMENTATION FOR BETTER RECOGNITION ACCURACY

Heinz Hofbauer¹ • Ehsaneddin Jalilian¹ • Andreas Uhl¹

¹Multimedia Signal Processing and Security Lab, University of Salzburg, Austria, {ejalilian, hofbauer, uhl}@cs.sbg.ac.at

Abstract

CNN-based iris segmentations have been proven to be superior to traditional iris segmentation techniques in terms of segmentation error metrics. To properly utilize them in a traditional biometric recognition systems requires a parameterization of the iris, based on the generated segmentation, to obtain the normalised iris texture typically used for feature extraction. This is an unsolved problem. We will introduce a method to parameterize CNN based segmentation, bridging the gap between CNN based segmentation and the rubbersheet-transform. The parameterization enables the CNN segmentation as full segmentation step in any regular iris biometric system, or alternatively the segmentation can be utilized as a noise mask for other segmentation methods. Both of these options will be evaluated.

Contents

1 Introduction	2
2 Related Work and State of the Art	2
2.1 The Iris Recognition Biometric System	2
2.2 CNNs in Iris Biometrics	3
3 Description of Algorithms	4
3.1 CNN-based Mask Generation	4
3.2 Parameterizing the Masks for Normalization	4
4 Experimental Evaluation	5
4.1 Impact of CNN Segmentation on the Recognition Rate	5
4.2 CNN Iris Detection as Replacement Noise Masks	7
5 Conclusion	8

1 Introduction

In past decades, iris recognition [1–3] has emerged as a rapidly growing field of research. Due to its intricate structure, the iris constitutes one of the most powerful biometric characteristics utilized by iris recognition algorithms to extract discriminative biometric reference data (templates). The ever-increasing demand on biometric systems has entailed continuous proposals of different iris recognition techniques [2]. Still, the processing chain of traditional iris recognition systems has remained almost unaltered. In particular, generic iris recognition systems consist of four major building blocks (see also Figure 2): (1) Iris image acquisition, (2) Image preprocessing, (3) Iris texture feature extraction, (4) Comparison (feature matching).

Recent methods utilizing convolutional neural networks (CNN) improve on different aspects of this pipeline. There are CNNs for segmentation of the iris, [4–7], and for feature extraction, [8, 9]. Some CNNs even replace the whole biometric toolchain and directly compare two iris recordings, [10].

CNN based segmentation of the iris, i.e., a binary mask separating iris from non-iris pixels, has shown great promise. In such an approach, a segmentation network is trained on ground-truth, manual iris segmentations, and evaluated on similar ground-truth augmented databases. The quality of segmentation is assessed based on type-1 and type-2 errors, as used in the noisy iris challenge evaluation (NICE), or the F-measure, i.e., the harmonic mean of recall and precision.

While these methods have shown good performance w.r.t. segmentation measures, they have not been evaluated as part of the biometric toolchain, i.e., not based on recognition performance. The reason for this is that CNNs generate a mask separating iris pixel from background pixels. However, the biometric toolchain as described above almost universally requires the transformation of the iris texture from polar coordinates (basically an unrolling of the iris), such that the angular dimension is mapped to the x-axis in a Cartesian coordinate system. This is done for easier storage and the orientation of radial features into a single direction, along the y-axis. Further, this ‘rubbersheet’ transformation stretches the iris texture to a fixed rectangular size which constitutes a normalisation step.

To the authors’ best knowledge, there are no methods to create a parameterization using the CNN based iris segmentation. We will introduce a method to parameterize the iris based on the CNN segmentation. While this is an easy task if the shape is closed and large parts of the iris are visible, it can also be challenging if there is a lot of obfuscation of the iris, e.g., by eyelids or eyelashes. An example of this is shown in Figure 1, where the upper iris image shows relatively little obfuscation resulting in a nice mask where the circular pupillary region can easily be identified and the outer iris border is also of high quality. In contrast, the lower image shows an example where the pupillary border is no longer fully visible and the outer iris border is largely cut off. The whole iris, not just the visible part, should be parameterized, usually by a circle or an ellipse, to generate a consistent representation during the ‘rubbersheet’ transform. If not done in a consistent fashion over different iris recordings, recognition performance will inevitably suffer [11]. A ‘noise’ mask can be used to mark the parts of the iris which

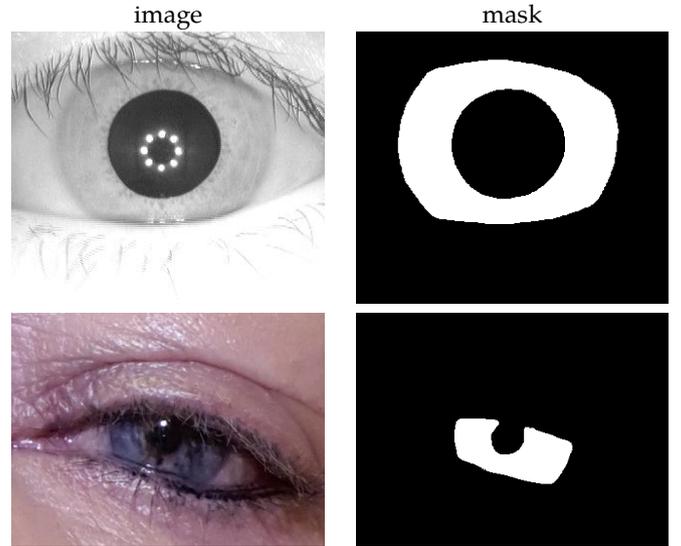


Figure 1: Examples of iris and mask which are easy and hard to parameterize for the rubbersheet transform.

are not visible, e.g., due to being overlapped by eyelids. These parts will then not be used during biometric comparison. We introduce our proposed method, and discuss the various steps to avoid problem cases, in Section 3.

An alternative is to use traditional segmentation methods, e.g., circular or elliptical parameterization, and then use the CNN based segmentation, interpreted as a pixel based iris/non-iris classification, as a noise mask. This noise mask is used during biometric comparison to omit non-iris areas, e.g., eyelids, eyelashes and reflections. As the masks produced by traditional segmentation methods are typically of relatively low quality, high quality CNN segmentations have the potential to improve the biometric recognition. This will be tested, along with the parameterization of the CNN segmentation, in Section 4.

2 Related Work and State of the Art

In this section, we will briefly describe the state of the art regarding the iris biometric toolchain and convolutional neural networks, specifically in the setting of iris biometrics.

2.1 The Iris Recognition Biometric System

A schematic view of the biometric toolchain is shown in Figure 2. While steps relevant to the presented work will be described in more detail, we refer the reader to [12] for a general overview. Modern iris recognition algorithms operate on normalised representations of the iris texture obtained by mapping the area between inner and outer iris boundaries $P, L : [0, 2\pi] \rightarrow [0, m] \times [0, n]$, from a polar coordinate system to a Cartesian one, i.e., to “Faberger” or “Rubbersheet” coordinates (using angle θ and pupil-to-limbic radial distance r) [1]. These “Rubbersheet” coordinates are Cartesian (+normalized) and the texture is “stretched” to fit into a normalized representation independent of pupillary dilation: $R(\theta, r) := (1 - r) \cdot P(\theta) + r \cdot L(\theta)$. This transformation is then

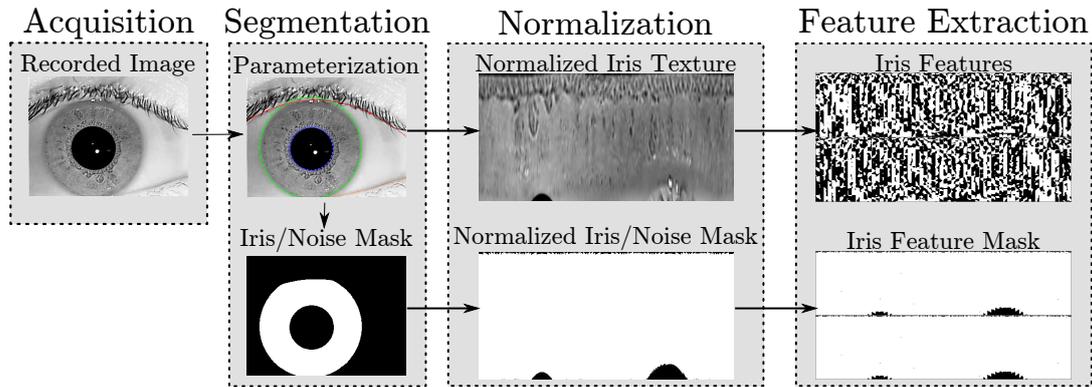


Figure 2: A schematic view of the biometric toolchain from image acquisition to matching.

used to generate a normalised image of the iris texture (R) and a noise masks (N). The latter usually considers reflections, upper and lower eyelid curves, and masks out occlusions caused by eyelashes such that $N(x, y) \neq 0$ if and only if pixel (x, y) refers to an in-iris location. While normalisation is standardised, there are several iris segmentation approaches for obtaining P , L and N . Traditional approaches employ circular boundary-based segmentation, such as Daugman’s integrodifferential operator [1] and Wildes’ circular Hough Transform (HT) [13]. Current iris segmentation techniques are often more involved, like the ellipsopolar transform for elliptical iris models [14].

2.2 CNNs in Iris Biometrics

Application of convolutional neural networks for iris segmentation has recently received attention in the research community, with different CNN based models proposed for this purpose [6]. Liu *et al.* [4] proposed a CNN classifier composed of three blocks of convolution and pooling layers, which takes in a local patch around pixels to extract features, and labels each pixel separately using a final fully connected layer (FC). However, this model lacks speed and efficiency. The authors proposed another model, which includes six blocks of interconnected convolution and pooling layers. The outs of the blocks then get fused together using a single fusion layer, which is flowed by a soft-max layer. However, performance evaluation is limited to a relatively small number of samples and does not go beyond segmentation, i.e., they did not assess performance in a recognition setup.

Jalilian and Uhl [6] used deep convolutional encoder-decoder networks in different variants [15], SegNet and derivatives, and trained them on annotated iris segmentation masks. The SegNet network comprises an encoder part and a decoder part and the ‘Basic’ and ‘Bayesian’ networks are just reduced versions of the same network to enable probabilistic pixel-wise segmentation using Monte-Carlo sampling and drop-out. In [7], the authors used the same ‘Basic’ network, and proposed a domain adaption model for iris images, eliminating the need for the annotated training masks.

Arsalan *et al.* [5] focus on the segmentation of visible light unconstrained (on-the-move) iris recordings. The CNN specifically focuses on the segmentation of the iris and requires a

prior ROI detection. An adapted version of a pre-trained VGG network is then used to identify iris pixel.

Arsalan *et al.* [16] propose an IrisDenseNet for iris end-to-end segmentation without preprocessing of the input image. The network has two main parts: a densely connected encoder and a SegNet decoder. While the network exhibits good performance, yet it is computationally intensive to train (specially when using bigger mini-batch sizes), due to dense connectivity.

Bazrafkan and Corcoran [17] propose a network for the segmentation task, in particular, a deep U-shaped network with 13 layers. It starts with a 3×3 kernel; the kernel size increases towards the center of the network to a maximum of 15×15 ; the kernel size then decreases towards the output. The network shows good performance and the authors found that the pooling operation results in unwanted artifacts at the output of the network.

Liu *et al.* [10] proposed a CNN classifier, called ‘DeepIris’, for heterogeneous iris verification, which learns relational features to measure the similarity between pairs of iris images. The network generates a similarity map between two images, and a similarity measure is calculated with help of a FC layer at the end. Minaee *et al.* [8] investigated the application of a VGG-style model as an iris feature extraction engine. They applied principal component analysis, and multi-class support vector machines on the extracted features to perform recognition.

Gangwar and Joshi [9] introduced two CNN classifiers for iris recognition (not for segmentation). The DeepIrisNet-A network is based on convolutional layers, and DeepIrisNet-B network uses a similar architecture, except using two inception layers after the convolutional layers. Both networks feed into a C-way softmax layer (where C is the number of classes) which produces a distribution over the class labels. Likewise, [18], proposed a smart-phone based iris recognition system using deep sparse filtering techniques to generate iris feature maps. None of the latter two works, however, do iris segmentation, and instead use the OSIRIS¹ algorithm for this purpose.

Zhao and Kumar [19] developed two fully convolutional networks (FCN) for feature extraction in the regular biometric toolchain. Specifically, they developed a feature extraction FCN, FeatNet, which takes a normalized iris texture as input

¹<http://svnext.it-sudparis.eu/svnview2-eph/ref-syst/Iris-0siris-v4.1>

CNN	Database	E1	E2	F1
RefineNet	<i>casia4i</i>	0.009	0.011	0.984
	<i>casiaA</i>	0.005	0.012	0.972
	<i>iitd</i>	0.015	0.018	0.974
	<i>protI</i>	0.013	0.036	0.823
iFCEDN[6]	<i>casia4i</i>	0.021	0.028	0.962
	<i>casiaA</i>	0.007	0.020	0.966
	<i>iitd</i>	0.018	0.022	0.970
	<i>protI</i>	0.051	0.030	0.947

Table 1: Average scores per dataset based on type 1 (E1) and type 2 (E2) errors as used in the noisy iris challenge evaluation (lower is better) and the F1-measure (F1, higher is better).

Parameter	Batchsize	Epoch	Momentum	Learning rate
Values	8	1000	0.1	0.9

Table 2: The training parameters for the network.

and generates an iris code which can be used in matching. In addition to that, they developed a second FCN, MaskNet, which also takes a normalized iris texture and generates a noise mask for use in matching.

3 Description of Algorithms

3.1 CNN-based Mask Generation

We considered RefineNet since it exhibits good performance for semantic segmentation (see experiments and comparison in [20]), which is an arguably harder task than iris segmentation (multiple versus two classes), and iFCEDN which was used in iris segmentation with good results in the past. Both have, in their respective publications, shown their superiority over a host of other architectures.

In this work, we use RefineNet [20], i.e., a multi-path refinement network, which employs a cascaded architecture with four RefineNet units, each of which directly connects to the output of one Residual net [21], as well as to the preceding RefineNet block in the cascade. Each RefineNet unit consists of two residual convolution units (RCU), whose outputs are fused into a high-resolution feature map, and then fed into a chained residual pooling block. The final part of each RefineNet block is another residual convolution unit.

RefineNet was chosen over other CNN architectures, like the improved FCEDN [6] (iFCEDN), for its higher segmentation performance, i.e., higher F-measure for generated masks, compared to a mask produced by human annotation (ground-truth). Table 1 lists a comparison of iFCEDN vs. RefineNet performance. For information about the ground-truth and data sets, we refer the reader to Section 4.

To assess segmentation performance, i.e., the generation of binary iris masks, we tested the CNN on all samples in each database, without overlapping training and testing sets, using five-fold cross-validation. To do so, we initially partitioned each database into five equal-sized subsets. Of the five subsets,

a single subset was retained as the testing data, and the remaining four subsets were used for training. The cross-validation process was then repeated five times, with each of the five subsets used exactly once as the testing data. In this way results for the whole database were generated without ever overlapping test and training data. Additional parameter settings relevant for training are listed in Table 2, the network architecture is unmodified from [20] as provided by the authors².

3.2 Parameterizing the Masks for Normalization

The parameterization process involves the following steps: (1) preprocessing by median blurring to smooth out the edges, (2) generation of candidate segmentations using a circular Hough transform [22], and (3) selecting the best candidate as a final parameterization. An example for the pupillary parameterization is shown in Figure 3. The annuli (see below) belonging to the chosen pupillary parameterization are also shown.

If the core parameters for the Hough transformation are set appropriately, i.e., pupillary radius no larger than 1/4th of the smallest image dimension and iris radius greater than pupillary radius, then the resulting circles almost certainly contain the correct parameterization. The selection process then is key to finding well fitting and stable parameterizations.

Given that the pupillary boundary is more complete than the iris boundary (simply because it is less obfuscated by eyelids and eyelashes), we start with that. The goal is to find the best parameterizations for which the parameterized circle fulfills the following properties: (a) surrounded by iris pixels, and (b) containing no iris pixels. We can use annuli adjacent to the circle candidate c , with center at c_x, c_y and radius c_r , and measure the number of iris pixels contained therein. Let the inner annulus be $A_i(c)$ with radii c_r and $0.5 \times c_r$, and likewise the outer annulus $A_o(c)$ with radii $(1.5 \times c_r, c_r)$. Let $N(a)$ count the number of iris pixels in area a and $\widehat{N}(a)$ be maximum possible number of iris pixels in area a . We can then use these (pupillary) value functions (normalized to $[0, 1]$) as a quality measure for each circle:

$$V^p(c) = \frac{N(A_o(c)) - N(A_i(c))}{N(\widehat{A}_o(c))}. \quad (1)$$

The circle candidate with the highest value will be the circle we choose as parameterization for the pupil c^p . If two candidates have the same value, the one with the smaller radius shall be chosen, in order to maximize the available iris texture.

The selection of the limbic circle is done in a similar manner, but changing the role of outer and inner annulus, since the limbic boundary should contain all the iris pixels. The largest radius candidate should be used in case of multiple candidates having the same value. Furthermore, there is an additional constraint which can be added to the selector function: the limbic and pupillary parameterizations should be roughly concentric. This leads to the limbic value function

²<https://github.com/guosheng/refinenet>

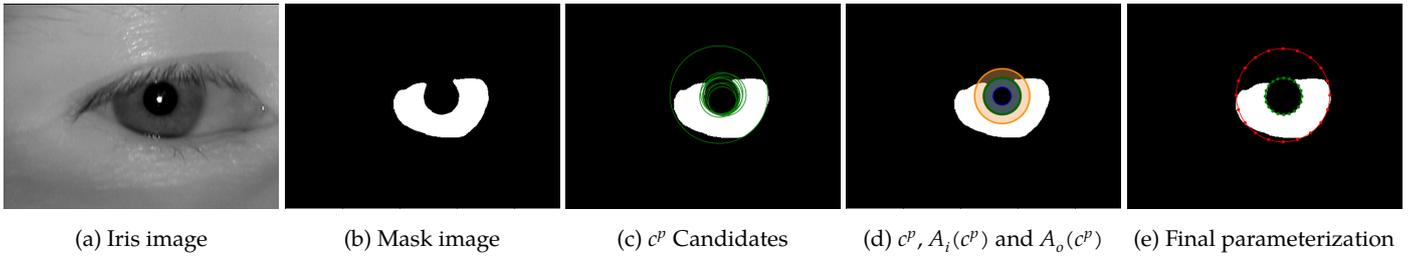


Figure 3: Outline of the parameterization process.

$$V^l(c, c^p) = \max\left(0, \frac{\text{border of iris texture term}}{N(\overline{A}_i(c)) - N(\overline{A}_o(c))} - \frac{\sqrt{(c_x - c_x^p)^2 + (c_y - c_y^p)^2}}{c_r}\right). \quad (2)$$

The lower cap of zero needs to be in place to keep it normalized in a $[0, 1]$ range, since, of course $N(\overline{A}_o) > N(\overline{A}_i)$. We further reduce the score by the relative radial offset. The cases where V^i would be less than zero are clearly not good parameterization candidates and as such the loss in differentiation does not impact the result. Like above, we choose the highest value circle c^l as the selected parameterization of the limbic boundary, and normalization is then done by transformation to ‘‘Faberge’’ coordinates in the regular fashion with $P := c^p$, $L := c^l$.

A few notes on the algorithm and parameter settings are worth pointing out. The first is to start with the pupillary boundary no matter which light source is used, near infrared or visible light. There are two reasons for this: i) the circular shape is more intact for the pupil which makes it easier to find a stable starting point and ii) the CNNs, at least over all the databases we looked at, do a very good job of generating an accurate pixel mask independent of the illumination wavelength. The second choice was to use a circular Hough transform instead of an elliptical curve fitting which is problematic for off-angle recordings. The main reason is that ellipses fitted to the iris masks would tend to be overly oblong due to occlusion of the iris by eyelids or eyelashes, see Figure 1 (lower right) for an example of this. To avoid abnormal ellipse formation would require more complicated value functions, which would be fine, but differences between strong off-angle recordings and minor occlusions would be almost impossible to detect. This, in turn, would lead to a more inconsistent parameterization which is detrimental to the overall process, see [11].

4 Experimental Evaluation

We used USIT 2 [23] as our main toolkit, since it is modular and allows to switch the segmentation without changing other parts of the system [24]. This allows for an analysis of only the change in segmentation. We used wavelet based (*qsw*) [25] and log-Gabor features (*lg*) [26]. For segmentation comparison, we used the circular (*CAHT*) [1] as well as the weighted adaptive ellipsopolar Hough transform (*WAHET*) [14]. Furthermore, we employed a method based on a total variation

model (*TVMIRIS*) [27] which also produces a parameterization³. The parameters used were those from the source code suggested for the UBIRIS database. These settings resulted in the best segmentations. The CNN-based segmentation with the parameterizations described in Section 3.2 will be denoted as *CNNHT*. The flexibility of the USIT, which allows to easily switch parts of the toolchain, has a cost, usually a slightly higher error rate, depending on the combination of algorithms [24]. To show this cost we also included a well known method from the literature, i.e., the OSIRIS [28] biometric toolchain.

The databases used are the well known IIT Delhi Iris Database version 1.0⁴ (*iitd*), the Interval subset of the CASIA Iris Image Database version 4.0⁵ (*casia4i*), and a subset of the ND-0405 Iris Image dataset (*ndi*)⁶ for which the segmentation ground-truth is available (837 iris images from 30 different subjects). In addition, we used a subset of the CASIA Iris Subject Ageing Version 1.0 Database (*casiaA*)⁷. The ground-truth for the segmentations is available for all three databases, for *iitd*, *casia4i* and *ndi* refer to [29]⁸ and for *casiaA* refer to [30]⁹.

Further, we included the PROTECT Multimodal DATABASE¹⁰ ([31]). Since this database is new we will give a brief overview. It contains different modalities, of which we only used the iris images (*protI*), from 47 subjects of great variety w.r.t. several aspects, including age and gender. The age interval is 21 to 76 and the distribution male/female is 57%/43%. The database is a visible light iris database at a distance with off angle acquisition and was included specifically because it is a very challenging database. The ground-truth is based on one human annotator with ellipse+polynomial contours, cf. [29]. The ground-truth of this database will be published jointly with this paper and will be made available at the Wavelab homepage¹¹.

4.1 Impact of CNN Segmentation on the Recognition Rate

Table 3 lists the results of this evaluation. The parameterization and the noise mask is produced by the given segmentation tool. The evaluation was performed on the full statistics of the

³available online at <http://www4.comp.polyu.edu.hk/~csajaykr/tvmiris.htm>

⁴http://www4.comp.polyu.edu.hk/~csajaykr/IITD/Database_Iris.htm

⁵<http://biometrics.idealtest.org/dbDetailForUser.do?id=4>

⁶http://www3.nd.edu/~cvrl/CVRL/Data_Sets.html

⁷<http://biometrics.idealtest.org/dbDetailForUser.do?id=14>

⁸<http://www.wavelab.at/sources/Hofbauer14b>

⁹<http://www.wavelab.at/sources/Hofbauer16d>

¹⁰<http://www.projectprotect.eu/database/>

¹¹<http://www.wavelab.at/sources/ProtI-GT>

Table 3: Biometric recognition performance, with segmentation masks, for the EER and FNMR@FMR=0.01% operation points. Segmentation and mask errors are also given.

Testset	Feat.	Seg.	EER	OP _{0.01}	SE	ME
<i>casiaA</i>	<i>lg</i>	OSIRIS	2.216	99.998	0	1406
<i>casiaA</i>	<i>lg</i>	TVMIRIS	22.036	99.975	13	20657
<i>casiaA</i>	<i>lg</i>	CAHT	4.786	12.009	0	85951
<i>casiaA</i>	<i>lg</i>	WAHET	13.899	99.563	0	0
<i>casiaA</i>	<i>lg</i>	CNNHT	2.613	13.770	0	0
<i>casiaA</i>	<i>lg</i>	Groundtruth	1.455	12.565	0	0
<i>casiaA</i>	<i>qsw</i>	OSIRIS	2.289	99.998	0	1406
<i>casiaA</i>	<i>qsw</i>	TVMIRIS	19.948	99.974	13	20657
<i>casiaA</i>	<i>qsw</i>	CAHT	4.636	8.617	0	85951
<i>casiaA</i>	<i>qsw</i>	WAHET	14.822	86.820	0	0
<i>casiaA</i>	<i>qsw</i>	CNNHT	2.685	9.451	0	0
<i>casiaA</i>	<i>qsw</i>	Groundtruth	1.310	5.761	0	0
<i>casia4i</i>	<i>lg</i>	OSIRIS	0.320	0.619	0	0
<i>casia4i</i>	<i>lg</i>	TVMIRIS	21.658	99.993	29	342080
<i>casia4i</i>	<i>lg</i>	CAHT	0.638	0.902	0	5277
<i>casia4i</i>	<i>lg</i>	WAHET	1.843	17.305	0	0
<i>casia4i</i>	<i>lg</i>	CNNHT	1.085	2.699	0	0
<i>casia4i</i>	<i>lg</i>	Groundtruth	0.244	0.414	0	0
<i>casia4i</i>	<i>qsw</i>	OSIRIS	0.352	0.515	0	0
<i>casia4i</i>	<i>qsw</i>	TVMIRIS	22.296	99.993	29	342080
<i>casia4i</i>	<i>qsw</i>	CAHT	0.694	0.837	0	5277
<i>casia4i</i>	<i>qsw</i>	WAHET	1.679	7.630	0	0
<i>casia4i</i>	<i>qsw</i>	CNNHT	1.284	3.141	0	0
<i>casia4i</i>	<i>qsw</i>	Groundtruth	0.213	0.224	0	0
<i>iitd</i>	<i>lg</i>	OSIRIS	1.629	24.974	0	0
<i>iitd</i>	<i>lg</i>	TVMIRIS	21.430	99.994	77	722481
<i>iitd</i>	<i>lg</i>	CAHT	1.308	20.726	0	71424
<i>iitd</i>	<i>lg</i>	WAHET	7.547	58.927	0	8956
<i>iitd</i>	<i>lg</i>	CNNHT	1.103	37.183	0	0
<i>iitd</i>	<i>lg</i>	Groundtruth	0.357	25.253	0	0
<i>iitd</i>	<i>qsw</i>	OSIRIS	1.521	26.503	0	0
<i>iitd</i>	<i>qsw</i>	TVMIRIS	23.028	99.994	77	722481
<i>iitd</i>	<i>qsw</i>	CAHT	1.173	19.491	0	71424
<i>iitd</i>	<i>qsw</i>	WAHET	7.783	47.782	0	8956
<i>iitd</i>	<i>qsw</i>	CNNHT	1.385	43.470	0	0
<i>iitd</i>	<i>qsw</i>	Groundtruth	0.305	29.068	0	0
<i>protI</i>	<i>lg</i>	OSIRIS	22.422	69.662	0	19980
<i>protI</i>	<i>lg</i>	TVMIRIS	25.929	58.076	2	0
<i>protI</i>	<i>lg</i>	CAHT	20.573	46.434	0	192071
<i>protI</i>	<i>lg</i>	WAHET	30.887	100.000	0	0
<i>protI</i>	<i>lg</i>	CNNHT	8.220	31.829	0	0
<i>protI</i>	<i>lg</i>	Groundtruth	7.899	29.565	0	0
<i>protI</i>	<i>qsw</i>	OSIRIS	21.698	62.473	0	19980
<i>protI</i>	<i>qsw</i>	TVMIRIS	26.556	59.056	2	0
<i>protI</i>	<i>qsw</i>	CAHT	23.022	47.359	0	192071
<i>protI</i>	<i>qsw</i>	WAHET	32.167	100.000	0	0
<i>protI</i>	<i>qsw</i>	CNNHT	9.701	32.692	0	0
<i>protI</i>	<i>qsw</i>	Groundtruth	9.309	30.466	0	0
<i>ndi</i>	<i>lg</i>	OSIRIS	25.466	39.272	0	0
<i>ndi</i>	<i>lg</i>	TVMIRIS	30.777	99.985	3	3199
<i>ndi</i>	<i>lg</i>	CAHT	27.046	46.340	0	47705
<i>ndi</i>	<i>lg</i>	WAHET	27.706	99.742	0	0
<i>ndi</i>	<i>lg</i>	CNNHT	24.703	40.758	0	0
<i>ndi</i>	<i>lg</i>	Groundtruth	24.212	36.923	0	0
<i>ndi</i>	<i>qsw</i>	OSIRIS	24.863	36.306	0	0
<i>ndi</i>	<i>qsw</i>	TVMIRIS	28.763	99.984	3	3199
<i>ndi</i>	<i>qsw</i>	CAHT	26.461	43.568	0	47705
<i>ndi</i>	<i>qsw</i>	WAHET	26.238	86.191	0	0
<i>ndi</i>	<i>qsw</i>	CNNHT	23.492	37.218	0	0
<i>ndi</i>	<i>qsw</i>	Groundtruth	23.906	33.557	0	0

database, that is all possible comparisons were performed. The results are given at two operation points of the ROC curve, i.e., the equal error rate (EER) and the false non match rate (FNMR) at a false match rate (FMR) of 0.01% (as OP_{0.01} in the table). All results listed in Table 3 are given in percentages. We further provide the number of errors produced by the system, of which there are two types. First, segmentation errors (SE) are errors where the segmentation fails to produce a usable parameterization. Second, masking errors (ME) are errors where the overlap of the two noise masks from the enrolled image and the probe presented to the system, leaves no pixel for comparison. Since only pixels not rejected by either mask are used for comparison, no score can be determined. It should be noted that, depending on system design, errors happen at different points in the system. Specifically, *CAHT* and *WAHET* are designed to be robust and will generally generate a parameterization. Failure in these cases is postponed to a mask error. The *CNNHT*, on the other hand, fails at the segmentation stage, since the parameterization used in the segmentation is directly based on the iris/noise mask.

It should be stated that the table also includes results from a ground-truth segmentation, i.e., a segmentation by a human observer who went over each image, delimited pupillary and iris boundaries and marked the upper and lower eyelid. This was done to assess the near optimal score which can be achieved given the texture quality of the database.

The first conclusion drawn from the results (Table 3) must be that the segmentation performance, in terms of segmentation and masking errors, of the CNNs in combination with the proposed parameterization algorithm, is very good. In fact, not a single error is produced, counting both segmentation and masking errors. Comparing the methods, we can see that *WAHET* produces less errors than *CAHT*, but still more than *CNNHT*. For *WAHET*, this comes at the expense of recognition performance and it shows the highest error rates over all feature extraction methods and databases. *OSIRIS* also produces no errors except for a low number of masking errors on *casiaA*.

The second conclusion is that the more complex *TVMIRIS* approach results in fewer errors (SE+ME) when it comes to the difficult test sets (*ndi* and *protI*), while maintaining a similar, but slightly worse, EER compared to *CAHT* and *WAHET*. However, this does not apply to all databases. For the difficult *casiaA*, it shows a similar rate of errors with a 5× higher EER compared to *CAHT*. Similarly, for high quality databases, *CAHT* and *WAHET* outperform *TVMIRIS* substantially with an error rate that is an order of magnitude lower.

The comparison of *CNNHT* and *CAHT* is a more interesting one. While *CAHT* produces a lot more errors, this could also be interpreted as rejecting low quality images. A masking error means the combined mask of both images does not allow for matching. In case of *casia4i*, a single image produced a failed segmentation, and consequently a noise mask rejecting every pixel. That image will then produce a masking error whenever it is used in a matching attempt. That this is a failure of the algorithm rather than the rejection of an image (which is of bad quality) is easily verified by looking at Figure 4. The figure also includes a sample from *protI* on which *CAHT* fails.



Figure 4: Sample images from the *casia4i* and *protI* databases on which *CAHT* failed to generate a proper segmentation.

The *CAHT* segmentation algorithm utilizes texture information as well as structure of the iris for segmentation and mask generation. With *CNNHT* the texture information is lost and can not be used during segmentation. For high quality images, in terms of biometric quality (i.e., open eyelids, clear iris texture and few obfuscations), this benefits the traditional segmentation algorithms. The less the texture can be used, the more similar the data during parameterization becomes between *CAHT*, *WAHET* and *CNNHT*. In such circumstances, the *CNNHT* has a better segmentation performance which in turn leads to a better parameterization. The results show this clearly: *casia4i* and *iitd* are high quality databases with frontal NIR images, and the results indicate that *CAHT* mostly outperforms *CNNHT* and *WAHET*. For the more difficult databases, i.e., *casiaA* and *ndi*, *CNNHT* starts to outperform *CAHT* as well. Importantly, on the *protI* database, which uses visible light and contains extremely difficult images, *CNNHT* segmentation shows great promise, with the next best segmentation more than doubling the EER. It should also be noted that the *CNNHT* almost reaches the performance of the ground-truth on this database (EER 7.9% versus 8.2%). The complexity and resulting performance of OSIRIS is between *CAHT/WAHET* on the one side and *CNNHT* on the other. On high quality databases (*iitd*, *casia4i*), it shows roughly the same performance as *CAHT*, on more difficult databases (*ndi* and *casiaA*) the performance is close to *CNNHT*, and on difficult databases (*protI*) the *CNNHT* is substantially better. This is further exemplified on the *ndi* database using *qsw* features, where the *CNNHT* (EER 23.492) actually improves on the *Groundtruth* (EER 23.906) by a statistically significant margin (assess by a McNemar test: $\chi^2 \approx 12.7$, $p^* \approx 0.04\%$).

Overall, the segmentation performance of the CNN and the proposed parameterization algorithm is excellent and the recognition performance is high. It is only surpassed on very high quality iris images where the traditional segmentation methods are expected to perform well.

4.2 CNN Iris Detection as Replacement Noise Masks

Noise masks are used to exclude parts of the normalised iris texture during biometric comparison which are occluded, e.g., by eyelids. Segmentation tools like *CAHT*, produce a basic noise mask and a parameterization of the iris. The more sophisticated CNN based iris/noise masks can be used as replacement noise masks by ‘normalizing’ them according to

the segmentation/parameterization. The expectation is that higher quality noise mask of the CNN will result in better biometric recognition.

Table 4: Biometric recognition performance, with the CNN-based mask instead of the original noise mask.

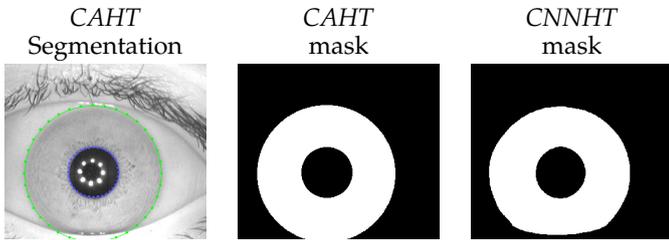
Testset	Feat.	Seg.	Δ EER	Δ OP _{0.01}	Δ ME
<i>casiaA</i>	<i>lg</i>	<i>TVMIRIS</i>	-2.935	+0.019	+239759
<i>casiaA</i>	<i>lg</i>	<i>CAHT</i>	-0.219	+87.979	-51890
<i>casiaA</i>	<i>lg</i>	<i>WAHET</i>	-1.51	+0.433	+22876
<i>casiaA</i>	<i>qsw</i>	<i>TVMIRIS</i>	-2.28	+0.019	+239759
<i>casiaA</i>	<i>qsw</i>	<i>CAHT</i>	-0.117	+91.373	-51890
<i>casiaA</i>	<i>qsw</i>	<i>WAHET</i>	-1.22	+13.159	+22876
<i>casia4i</i>	<i>lg</i>	<i>TVMIRIS</i>	-9.464	-68.37	-336861
<i>casia4i</i>	<i>lg</i>	<i>CAHT</i>	+ 0.023	+0.07	-5277
<i>casia4i</i>	<i>lg</i>	<i>WAHET</i>	-0.444	-15.328	0
<i>casia4i</i>	<i>qsw</i>	<i>TVMIRIS</i>	-8.799	-71.067	-336861
<i>casia4i</i>	<i>qsw</i>	<i>CAHT</i>	-0.002	+0.059	-5277
<i>casia4i</i>	<i>qsw</i>	<i>WAHET</i>	-0.252	-5.773	+0
<i>iitd</i>	<i>lg</i>	<i>TVMIRIS</i>	-13.205	-68.125	-718156
<i>iitd</i>	<i>lg</i>	<i>CAHT</i>	+ 0.368	+10.769	-66915
<i>iitd</i>	<i>lg</i>	<i>WAHET</i>	-1.356	+40.495	-4425
<i>iitd</i>	<i>qsw</i>	<i>TVMIRIS</i>	-14.203	-65.663	-718156
<i>iitd</i>	<i>qsw</i>	<i>CAHT</i>	+0.415	+9.197	-66915
<i>iitd</i>	<i>qsw</i>	<i>WAHET</i>	-1.557	+31.717	-4425
<i>protI</i>	<i>lg</i>	<i>TVMIRIS</i>	-2.32	+41.913	4005
<i>protI</i>	<i>lg</i>	<i>CAHT</i>	+11.972	+53.556	-185599
<i>protI</i>	<i>lg</i>	<i>WAHET</i>	+0.771	-0.002	+11093
<i>protI</i>	<i>qsw</i>	<i>TVMIRIS</i>	-1.258	+40.93	+4005
<i>protI</i>	<i>qsw</i>	<i>CAHT</i>	+10.078	+52.631	-185599
<i>protI</i>	<i>qsw</i>	<i>WAHET</i>	+1.059	-0.002	+11093
<i>ndi</i>	<i>lg</i>	<i>TVMIRIS</i>	-1.908	+0	+17011
<i>ndi</i>	<i>lg</i>	<i>CAHT</i>	+0.8	+9.433	-47701
<i>ndi</i>	<i>lg</i>	<i>WAHET</i>	-2.42	-38.427	0
<i>ndi</i>	<i>qsw</i>	<i>TVMIRIS</i>	-1.166	+0	+17011
<i>ndi</i>	<i>qsw</i>	<i>CAHT</i>	+0.502	+8.924	-47701
<i>ndi</i>	<i>qsw</i>	<i>WAHET</i>	-1.932	-28.36	0

Table 4 gives the change in EER, OP_{0.01} and ME compared to Table 3 when the CNN-based mask is used instead of the mask produced by the actual segmentation. OSIRIS does not output a parameterization and as such the CNN noise mask could not be interjected into the toolchain.

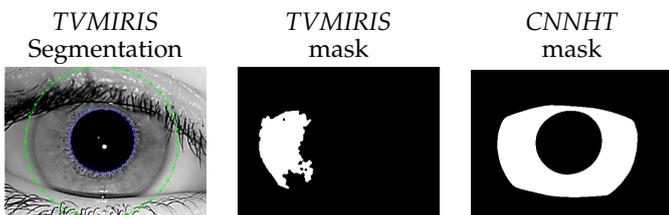
Overall, the use of the *CNNHT* as a drop-in mask is possible, but the results are somewhat mixed. The improvement is small, see *casiaA*, since the mask only reduces comparison errors due to obfuscations and can not correct the actual segmentation, which is of prime importance [11]. On less difficult databases, the masks produced by the regular algorithms are sufficient and only minor improvements happen, but not on a consistent basis. Likewise, for difficult databases, the change of masks does not help at all since the segmentation is inconsistent. Examples of this can be seen in Figure 5. In particular, Figure 5a shows an example where the segmentation of *CAHT* is simply wrong which is not changed by a mask centered



(a) Sample from the *protI* database.



(b) Sample from the *casia4i* database.



(c) Sample from the *iitd* database.

Figure 5: Comparison of CAHT and TVMIRIS segmentation and mask with CNNHT mask.

around a different part of the image. Figure 5b, on the other hand, shows an example of a high quality iris image and its corresponding segmentation where both masks are basically the same. In this second case, the CNNHT mask is slightly better since it correctly includes the lower eyelid, but the difference is marginal. This leads to the occasional marginal improvements seen in Table 4. Figure 5c shows that true improvement happens (TVMIRIS on *iitd* EER from 21% to 8%) when the original algorithm has errors in the mask generation, or is too conservative. In both cases, the CNNHT can supply a better mask.

5 Conclusion

We introduced a parameterization method which works well and, over all tested databases, produced no failures (to parameterize).

On high quality databases, specifically NIR images with open eyelids and frontal acquisition, the traditional iris segmentation methods outperform the CNN segmentation plus parameterization. However, on slightly lower quality databases, the CNN segmentation and parameterization outperform all other methods tested.

The use of CNN based segmentation as a noise mask for traditional, parameter based, segmentations leads to an improvement in a large number of cases. However, two cases clearly do

not show an improvement. First, when the biometric recording is of high quality, the CNN based noise mask shows no improvement. This is due to the low amount of occlusion to begin with. Second, on difficult data, the traditional methods often failed to properly segment the iris. In this case, the correct mask of the CNN based segmentation could not compensate the parts missed by the traditional methods.

Overall, CNN based segmentation and the proposed parameterization improves over traditional segmentation methods, except on very high quality biometric recordings. Using the CNN based segmentation as a noise mask improves the traditional methods, except on very high quality biometric recordings (where the original noise mask was sufficient), or very difficult recordings (where the traditional segmentations methods fail).

Acknowledgements

This project has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 700259.

References

- [1] J. Daugman, “How iris recognition works,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, 2004 (cit. on pp. 2, 3, 5).
- [2] K. Bowyer, K. Hollingsworth, and P. Flinn, “Image understanding for iris biometrics: A survey,” *Computer Vision and Image Understanding*, vol. 110, no. 2, 2008 (cit. on p. 2).
- [3] C. Rathgeb, A. Uhl, and P. Wild, *Iris Recognition: From Segmentation to Template Security*, ser. Advances in Information Security. 2013, vol. 59 (cit. on p. 2).
- [4] N. Liu, H. Li, M. Zhang, J. Liu, Z. Sun, and T. Tan, “Accurate iris segmentation in non-cooperative environments using fully convolutional networks,” in *2016 International Conference on Biometrics (ICB)*, 2016 (cit. on pp. 2, 3).
- [5] M. Arsalan, H. G. Hong, R. A. Naqvi, M. B. Lee, M. C. Kim, D. S. Kim, C. S. Kim, and K. R. Park, “Deep learning-based iris segmentation for iris recognition in visible light environment,” *Symmetry*, vol. 9, no. 11, 2017. doi: [10.3390/sym9110263](https://doi.org/10.3390/sym9110263) (cit. on pp. 2, 3).
- [6] E. Jalilian and A. Uhl, “Iris segmentation using fully convolutional encoder–decoder networks,” in *Deep Learning for Biometrics*, 2017, ch. 6, ISBN: 978-3-319-61656-8. doi: https://link.springer.com/chapter/10.1007/978-3-319-61657-5_6 (cit. on pp. 2–4).
- [7] E. Jalilian, A. Uhl, and R. Kwitt, “Domain adaptation for cnn based iris segmentation,” in *Proceedings of the IEEE 16th International Conference of the Biometrics Special Interest Group (BIOSIG’17)*, 2017. doi: <https://ieeexplore.ieee.org/document/8053502> (cit. on pp. 2, 3).
- [8] S. Minaee, A. Abdolrashidiy, and Y. Wang, “An experimental study of deep convolutional features for iris recognition,” in *2016 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, ser. SPIE Proceedings, 2016 (cit. on pp. 2, 3).

- [9] A. Gangwar and A. Joshi, "Deepirisnet: Deep iris representation with applications in iris recognition and cross-sensor iris recognition," in *2016 IEEE International Conference on Image Processing (ICIP)*, ser. SPIE Proceedings, 2016 (cit. on pp. 2, 3).
- [10] N. Liu, M. Zhang, H. Li, Z. Sun, and T. Tan, "Deepiris: Learning pairwise filter bank for heterogeneous iris verification," *Pattern Recognition Letters*, vol. 82, no. 2, 2016 (cit. on pp. 2, 3).
- [11] H. Hofbauer, F. Alonso-Fernandez, J. Bigun, and A. Uhl, "Experimental analysis regarding the influence of iris segmentation on the recognition rate," *IET Biometrics*, vol. 5, no. 3, 2016. doi: [10.1049/iet-bmt.2015.0069](https://doi.org/10.1049/iet-bmt.2015.0069) (cit. on pp. 2, 5, 7).
- [12] *Handbook of Iris Recognition*. 2016. doi: <https://doi.org/10.1007/978-1-4471-6784-6> (cit. on p. 2).
- [13] R. P. Wildes, "Iris recognition: an emerging biometric technology," in *Proceedings of the IEEE*, vol. 85, 1997 (cit. on p. 3).
- [14] A. Uhl and P. Wild, "Weighted adaptive hough and ellipsoidal transforms for real-time iris segmentation," in *Proceedings of the 5th IAPR/IEEE International Conference on Biometrics (ICB'12)*, 2012 (cit. on pp. 3, 5).
- [15] 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*, vol. PP, 99 2017 (cit. on p. 3).
- [16] M. Arsalan, R. A. Naqvi, D. S. Kim, P. H. Nguyen, M. Owais, and K. R. Park, "Irisdensenet: Robust iris segmentation using densely connected fully convolutional networks in the images by visible light and near-infrared light camera sensors," *Sensors*, vol. 18, no. 5, 2018, issn: 1424-8220. doi: [10.3390/s18051501](https://doi.org/10.3390/s18051501) (cit. on p. 3).
- [17] S. Bazrafkan and P. Corcoran, "Enhancing iris authentication on handheld devices using deep learning derived segmentation techniques," in *IEEE International Conference on Consumer Electronics, ICCE 2018*, 2018. doi: [10.1109/ICCE.2018.8326219](https://doi.org/10.1109/ICCE.2018.8326219) (cit. on p. 3).
- [18] K. Raja, R. Raghavendra, V. K. Vemuri, and C. Busch, "Smartphone based visible iris recognition using deep sparse filtering," *Pattern Recognition Letters*, vol. 57, no. Supplement C, 2015 (cit. on p. 3).
- [19] Z. Zhao and A. Kumar, "Towards more accurate iris recognition using deeply learned spatially corresponding features," in *The IEEE International Conference on Computer Vision (ICCV)*, 2017 (cit. on p. 3).
- [20] G. Lin, M. Anton, S. Chunhua, and I. Reid, "Refinenet: Multi-path refinement networks for high-resolution semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017 (cit. on p. 4).
- [21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016 (cit. on p. 4).
- [22] H. K. Yuen, J. Princen, J. Illingworth, and J. Kittler, "A comparative study of hough transform methods for circle finding," *Image Vision Computing*, vol. 8, no. 1, 1990 (cit. on p. 4).
- [23] University of Salzburg, *USIT – university of salzburg iris toolkit*, <http://www.wavelab.at/sources/USIT>, version 2.x, University of Salzburg, 2017 (cit. on p. 5).
- [24] C. Rathgeb, A. Uhl, P. Wild, and H. Hofbauer, "Design decisions for an iris recognition sdk," in *Handbook of Iris Recognition*, ser. Advances in Computer Vision and Pattern Recognition, second edition, 2016. doi: [10.1007/978-1-4471-6784-6](https://doi.org/10.1007/978-1-4471-6784-6) (cit. on p. 5).
- [25] L. Ma, T. Tan, Y. Wang, and D. Zhang, "Efficient iris recognition by characterizing key local variations," *IEEE Transactions on Image Processing*, vol. 13, no. 6, 2004, issn: 1057-7149. doi: [10.1109/TIP.2004.827237](https://doi.org/10.1109/TIP.2004.827237) (cit. on p. 5).
- [26] L. Masek, "Recognition of human iris patterns for biometric identification," Master's thesis, University of Western Australia, 2003 (cit. on p. 5).
- [27] Z. Zhao and A. Kumar, "An accurate iris segmentation framework under relaxed imaging constraints using total variation model," in *The IEEE International Conference on Computer Vision (ICCV)*, 2016 (cit. on p. 5).
- [28] D. Petrovska and A. Mayoue, "Description and documentation of the biosecure software library," Project No IST-2002-507634 - BioSecure, Deliverable, 2007 (cit. on p. 5).
- [29] H. Hofbauer, F. Alonso-Fernandez, P. Wild, J. Bigun, and A. Uhl, "A ground truth for iris segmentation," in *Proceedings of the 22th International Conference on Pattern Recognition (ICPR'14)*, (cit. on p. 5).
- [30] H. Hofbauer, I. Tomeo-Reyes, and A. Uhl, "Isolating iris template ageing in a semi-controlled environment," in *Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG'16)*, 2016 (cit. on p. 5).
- [31] U. of Reading, *PROTECT multimodal DB dataset*, Available by request at projectprotect.eu/dataset, 2017 (cit. on p. 5).