Deep Representations for Cross-spectral Ocular Biometrics

Luiz A. Zanlorensi[1], Diego R. Lucio[1]
Alceu S. Britto Jr.[2], Hugo Proença[3], David Menotti[1]

[1]Department of Informatics, Federal University of Paraná, Curitiba, Brazil
[2]Postgraduate Program in Informatics, Pontifical Catholic University of Paraná, Curitiba, Brazil
[3]IT: Instituto de Telecomunicações, University of Beira Interior, Covilhã, Portugal

[1]{*lazjunior, drlucio, menotti*}@inf.ufpr.br   [2]*alceu.junior@pucpr.br*   [3]*hugomcp@di.ubi.pt*

ABSTRACT

One of the major challenges in ocular biometrics is the cross-spectral scenario, i.e., how to match images acquired in different wavelengths (typically visible (VIS) *against near-infrared* (NIR)*). This article designs and extensively evaluates cross-spectral ocular verification methods, for both the closed and open-world settings, using well known deep learning representations based on the iris and periocular regions. Using as inputs the bounding boxes of non-normalized iris/periocular regions, we fine-tune Convolutional Neural Network* (CNN) *models (based either on VGG16 or ResNet-50 architectures), originally trained for face recognition. Based on the experiments carried out in two publicly available cross-spectral ocular databases, we report results for intra-spectral and cross-spectral scenarios, with the best performance being observed when fusing* ResNet-50 *deep representations from both the periocular and iris regions. When compared to the state-of-the-art, we observed that the proposed solution consistently reduces the Equal Error Rate* (EER) *values by* 90% / 93% / 96% *and* 61% / 77% / 83% *on the cross-spectral scenario and in the PolyU Bi-spectral and Cross-eye-cross-spectral datasets. Lastly, we evaluate the effect that the "deepness" factor of feature representations has in recognition effectiveness, and - based on a subjective analysis of the most problematic pairwise comparisons - we point out further directions for this field of research.*

1. INTRODUCTION

Iris recognition using near-infrared (NIR) wavelength images acquired under controlled environments can be considered a mature technology, which proved to be effective in different scenarios [1]. In contrast, performing iris recognition in uncontrolled environments and at visible (VIS) wavelength is still a challenging problem [2, 3]. Some of the latest researches consist of biometrics recognition on cross-spectral scenarios, i.e., using images of eyes from the same subject obtained at the VIS and NIR wavelengths [4–7].

Recently, machine learning techniques based on deep learning have been achieving great popularity due to the results reported in the literature, which advance the state-of-the-art in various problems, such as speech recognition [8–10], natural language processing [11, 12], digit and character recognition [13–15] and face recognition [16, 17]. In the field of ocular biometrics, using deep learning representation has been advocated both for the periocular [18, 19]

and iris [6, 20–26] regions, with interesting and promising results being reported.

As stated in previous works [20, 27], an often and open problem in ocular recognition is the matching heterogeneous images captured at different resolutions, distances and devices (cross-sensor and cross-spectral). Regarding these problems it is difficult to design a robust handcrafted feature extractor to address the intra-class variations present in this scenarios. In this sense, several recent works demonstrate that deep representations report better results compared to handcrafted features in iris and periocular region recognition [18–20, 25].

Having in mind that deep learning frameworks are typically able to produce robust representations, in this article we apply this family of frameworks to extract and combine features from the ocular region, obtained at different wavelengths, e.g., VIS and NIR. The strategy described in this article is composed of some methodologies extracted from the literature. For both the iris and ocular traits we use as input the bounding box delimited regions used in the state-of-the-art methods [18, 26]. Then, the features from these traits were extracted using a similar approach proposed by [26]. In this direction, the main contribution of this article is the extensive experiments on two datasets comparing iris, periocular, and fusion results for both cross-spectral (VIS to NIR) and intra-spectral (VIS to VIS, NIR to NIR) matching, reaching a new state-of-the-art results. There is also the following four-fold contributions: (i) we show that deep learning yield robust representations on two well-known cross-spectral databases (PolyU and Cross-Eyed) for ocular verification using closed- and open-world protocols; (ii) we report how two off-the-shelf networks can be fine-tuned from the face domain to the periocular and iris one; (iii) we analyze the use of a single deep representation extraction schema, for both cross-spectral and the same spectra scenarios; and (iv) we conclude about the benefits of fusing the periocular and iris representations to improve the recognition accuracy.

The remainder of this work is organized as follows. In Section 2, we describe some recent works that use deep learning for iris and periocular recognition. Section 3 provides the details of the proposed approach. Section 4 presents the databases, metrics and evaluation protocol used in our empirical evaluation. The results are presented and discussed in Section 5. Lastly, the conclusions are given in Section 6.

2. RELATED WORK

This section surveys the works that use deep learning frameworks for iris and periocular recognition. Also, we summarize the most relevant ocular recognition methodologies focused on the cross-spectral scenario.

One of the first works applying deep learning to iris recognition only was the *DeepIris* framework, proposed by Liu et al. [20]. Having as a goal the recognition of heterogeneous irises using images obtained by different sensors (i.e., the cross-sensor scenario), the authors proposed a framework that establishes the similarity between a pair of iris images using Convolutional Neural Networks (CNNs) by learning a bank of pairwise filters. The experiments were performed in the Q-FIRE and CASIA cross-sensor databases, reporting promising results with Equal Error Rate (EER) of 0.15% and 0.31%, respectively.

Another deep learning application for cross-sensor iris recognition, designated DeepIrisNet, was proposed by Gangwar & Joshi [21]. In their study, two CNN architectures were presented and used to extract features and representations of iris images. Comparing to the baselines, their methodology showed better robustness with respect to five different factors: effect of segmentation, image rotation, input size, training size, and network size.

Nguyen et al. [23] argued that generic descriptors yielding from deep learning frameworks can appropriately represent iris features from NIR images obtained in controlled environments. The authors compared five CNN architectures trained in the ImageNet database [28]: AlexNet, VGG, Inception, ResNet and DenseNet. Deep representations were extracted from normalized iris images at different depths of each CNN model. Afterward, a simple multi-class Support Vector Machine (SVM) was applied to perform the identification. The experiments were carried out in the LG2200 (ND-CrossSensor-Iris-2013) and CASIA-Iris-Thousand databases and compared with a baseline feature descriptor [29]. As main result, the authors argued that features extracted from intermediate layers of the networks reported better results than the representations in the deeper layers.

Luz et al. [18] extracted deep representations of the periocular region using the VGG16 CNN model. The authors reported promising results by using transfer learning techniques from the face recognition domain, followed by fine-tuning using the ocular images. The experiments achieved the state-of-the-art in the NICE.II and Mob-BIO databases, which were obtained in uncontrolled environments at the VIS wavelength.

Also using the NICE.II database, Silva et al. [30] proposed a fusion method of iris and periocular deep representations by means of feature selection using the Particle Swarm Optimization (PSO). Similar to the methodology proposed in [18], the iris and periocular deep representations were extracted with the VGG16 model trained for face recognition and fine-tuned for each trait. Promising results were reported in the verification mode only using iris information and also using iris and periocular fusion.

Proença and Neves [19] argue that periocular recognition performance is optimized when the iris and sclera regions are discarded. Also, these authors describe a processing chain based on CNN that defines the regions-of-interest in the input image. In their approach, a segmentation process is only required to create the training samples. This process consists in generating a periocular image of a subject containing an ocular (sclera and iris) region belonging to other subjects. Then, the generated samples are used for data augmentation and to feed the learning phase of the CNN model. The experiments were performed in the UBIRIS.v2 and FRGC databases and

consistently advances the state-of-the-art in the closed-world setting.

Zanlorensi et al. [26] evaluated the impact of the segmentation for noise removal and normalization when deep representations were extracted from the iris images. The experiments reported that deep representations extracted from an iris bounding box without segmentation process achieved better results than normalized and segmented images. In addition, the authors compared representations extracted from the VGG16 and ResNet50 models and the impact of using data augmentation techniques. A new state-of-the-art was reached in the NICE.II database using only information from the iris region.

In terms of cross-sensor iris recognition, the methodology proposed by Nalla and Kumar [6] introduced a domain adaptation framework to address this problem and reported a new approach using Markov random fields. The experiments were performed using two cross-sensor iris databases: IIT-D CLI and ND Cross sensor 2012; and one cross-spectral iris database: PolyU. The results reported in PolyU database in the verification protocol at closed-world achieved an EER value of 3.97% in NIR *vs* NIR comparisons and 6.56% in VIS *vs* VIS comparisons. Using the Markov random fields on cross-spectral comparisons, their methodology achieved 23.87% of EER.

In [25], the authors evaluated a range of deep learning architectures applied to the cross-spectral iris recognition. The experimental results were performed in the PolyU and Cross-Eyed databases. Experimental analysis indicates that iris features extracted from CNN models are generally sparse and can be used for template compression. Several hashing algorithms were evaluated and the most effective was supervised discrete hashing achieving more accurate performance and reducing the size of iris template. The best results reported were achieved by incorporating supervised discrete hashing on the deep representations extracted with a CNN model trained with a softmax cross-entropy loss. This methodology reached an EER value of 12.41% and 6.34% on the PolyU and Cross-Eyed databases, respectively. However, the authors do not report the system performance on the open-world protocol, which is a more realistic scenario. Also, this methodology requires an approach for the segmentation and normalization of the iris. To the best of our knowledge, this work is the state-of-the-art on cross-spectral recognition in the verification mode. Thus, it is used for comparison with the methodology presented in this paper.

Also applied in the cross-spectral scenario, Hernandez-Diaz et al. [31] proposed a method using a ResNet-101 model pretrained in the Imagenet database [28] to extract deep representation from periocular images. The experiments were carried out in verification mode using the IIITD Multispectral Periocular database [5] in three different spectra: Visible, Night Vision, and Near-Infrared. The results were reported using features extracted at each layer from the model using chi-square distance and cosine similitude to perform the matching. The authors stated that the features extracted from the intermediate layer from the ResNet-101 model achieved the best results in the cross-spectral experiments.

Recently, two contests were performed using the Cross-Eyed database, aiming to recognize iris and periocular (without the iris region) traits in a cross-spectral environment [32, 33]. However, as stated by Wang and Kumar [25], the results reported in these competitions should be considered preliminary, as they employed a comparison protocol with less matching challenge than usual (only 3 images of each class were used in the inter-class comparison instead of all against all) and did not provide information regarding which images of each class were used in the inter-class matching (the authors' work only reported that the images were randomly selected). Other problems include the availability of codes and also details of

the methodologies, which limit the reproducibility.

Previous works on cross-spectral recognition such as [6, 25] only use iris traits and require a methodology for iris segmentation and normalization. Our proposal in this article combine information from the iris and periocular regions. Also, for the iris trait, we use only a bounding box, which does not require segmentation for noise removal and normalization steps.

For completeness, there are several other applications with ocular images based on deep learning such as: spoofing detection [34], recognition of mislabeled left and right iris images [35], liveness detection [36], iris/periocular region location/detection [37, 38], sclera and iris segmentation [39, 40], gender classification [41] and sensor model identification [42].

3. METHODOLOGY

In this paper we analyze the use of deep representations from the eye regions (iris and periocular) on cross-spectral scenario, i.e., obtaining models able to match VIS against NIR wavelength images. Particularly, we evaluate and combine deep representations extracted from two modalities (traits): the iris and periocular regions. In the periocular modality, features were extracted from the entire image (considering the iris, sclera, skin, eyelids and eyelashes components). On the other way, the iris features were extracted from a bounding box, i.e., a cropped image that contains only the iris region, as described by Zanlorensi et al. [26]. These bounding boxes were generated manually by coarse annotations and are publicly available to the research community[1] and appears in [38]. Samples of the periocular and iris images used in this work are shown in Figure 1.
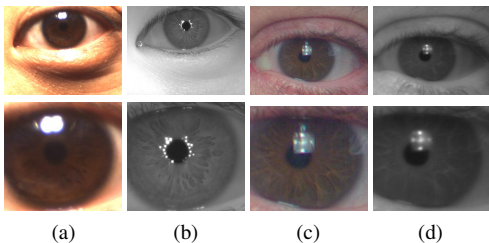

| (a) | (b) | (c) | (d) |

Fig. 1. VIS (a,c) and NIR (b,d) samples from the PolyU (a,b) and Cross-Eyed (c,d) databases. First and second rows show periocular and iris images, respectively.

Deep representations from the periocular and iris regions were extracted using a similar approach proposed in [26]. In this way, the VGG16 [16] and ResNet-50 [17] CNN models trained for face recognition were fine-tuned to each modality. We choose these models because they reported promising results in recent works applied in ocular recognition [18, 25, 26, 30]. The architecture modifications for both models consist of the removal of the last layer and the addition of two new layers. The first one is a fully-connected layer with 256 neurons that will be used as the feature representation and aim to reduce the feature dimensionality, since originally VGG16 and ResNet-50 have 4096 and 2048 features/outputs, respectively. The other layer added has a softmax cross-entropy loss function and it is used only in the training phase in an identification mode. We chose a feature vector of 256 features based on the results reported by Luz et al. [18], where the authors evaluated different feature vector sizes and stated that vector with such size (256) showed the best trade-off

regarding matching time, amount of memory required and matching effectiveness. The strategy applied to extract features from NIR and VIS images is detailed in Figure 2.
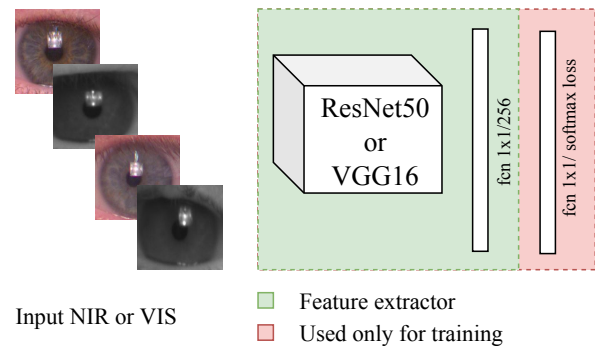

Fig. 2. The cross-/intra-spectral ocular recognition strategy. A single model (ResNet50 or VGG16) is used to learn features from both spectra: NIR and VIS.

The number of epochs used for training was chosen based on a validation subset composed of 20% of the training set images. After defining the number of epochs, the CNN models were trained using the entire training set. The training was performed with the Stochastic Gradient Descent (SGD) optimizer and without freezing any weights of the pre-trained layers.

In the test phase, as previously mentioned, the last layer of each model was removed and the features were extracted from the first new last layer, composed by 256 neurons.

The all-against-all matching was performed using the cosine distance metric, which measures the cosine of the angle between two vectors. Regarding the similarity of biometrics features/representations, it is known that orientation is more important than the magnitude coefficient. The cosine distance metric faithfully matches this feature, being given by:

$$d_c(A, B) = 1 - \frac{\sum_{j=1}^{N} A_j B_j}{\sqrt{\sum_{j=1}^{N} A_j^2} \sqrt{\sum_{j=1}^{N} B_j^2}}, \qquad (1)$$

where A and B stand for the feature vectors.

The iris and periocular region representations were combined, applying the score-level fusion technique. Similar to approaches that also used score-level fusion for iris and periocular region traits [6, 43, 44] and also based on the individual performance of each trait in our experiments, we chose to use weights of 0.6 and 0.4 for the periocular region and iris representations, respectively. To perform fusion at the score-level, first, we compute the matching for each trait independently, and then we calculated the weighted arithmetic mean between the cosine distances computed for the iris and periocular modalities.

It is important to note that, in the model learning process, all images (NIR and VIS) were used to feed the CNN models, making a single model to learn discriminant features of images captured in both spectra. To the best of our knowledge, this procedure is similar to the adopted in [25] for the CNN architecture. In the test phase the features are extracted for all images NIR or VIS images. However, note that for evaluating the cross-spectral scenario, only images acquired under different wavelengths are paired to match.

4. DATABASES, METRICS AND PROTOCOL

This section describes the databases used, the experimental protocol defined and the metrics considered appropriate to provide a meaningful comparison between our method and the baselines.

4.1. Databases

Two well-known databases were used in our empirical evaluation: 1) the PolyU; and the 2) Cross-Eyed databases, described below:

4.1.1. PolyU database

PolyU (PolyU Bi-spectral) database is composed of images obtained simultaneously under both NIR and VIS wavelengths. The entire database has 12,540 images with a resolution of 640×480 pixels. For every spectrum, there are 15 samples of each eye (left and right) from 209 subjects (418 classes) [6].

4.1.2. Cross-Eyed database

The Cross-Eyed (Cross-eyed-cross-spectral) iris database has 3,840 images from 120 subjects (240 classes). There are 8 samples from each of the classes for every spectrum. The resolution of the images is 400×300 pixels. All images were obtained at a distance of 1.5 meters, in an uncontrolled indoor environment, with a wide variation of ethnicity and eye colors, and lightning reflexes [32].

4.2. Metrics

For evaluating the algorithms, **we choose the EER metric**, which is determined by the intersection point of False Acceptance Rate (FAR) and False Rejection Rate (FRR) curves generated when the acceptation/rejection threshold is varied.

We also report the decidability score d' [45]. The metric or index d' measures how well separated are the two types of distributions (*genuines* and *impostors*), in the sense that recognition errors correspond to the regions where both distributions overlap:

$$d' = \frac{|\mu_E - \mu_I|}{\sqrt{\frac{1}{2}(\sigma_E^2 + \sigma_I^2)}}, \tag{2}$$

where the means and standard deviations of the genuine and impostor distributions are given by μ_I, μ_E, σ_I, and σ_E, respectively.

Whereas the index d' can be related to the feature vector discrimination ability of an approach, the EER metric measures the real performance of a biometric system. Therefore, regarding a real-world application, we consider the EER as the primary metric in the results reported in this work.

4.3. Protocol

In all experiments, the *verification* setting was the unique considered, in which pairs of images are compared in order to determine whether a subject is who he claims to be or not. For this, following a *one-against-all* pairwise matching strategy, all pairs of genuine and impostor comparisons were generated.

For a fair comparison with the state-of-the-art methods, the test protocol used in this work follows the procedures given in[6, 25], which consists of a *closed-world* protocol, where different instances of the same class are distributed in the training and test sets. In the PolyU database, the first ten instances from every subject were used for training and the remainder (five) were employed for the

matching. In the Cross-Eyed database, the first five instances from every subject are used for training and the remaining three instances were employed for the matching.

To perform the experiments, we considered that in both databases, the NIR and VIS images were obtained synchronously. Thus, here in the intra-class comparison in the cross-spectral scenario, images of the same index were not matched, because the pair represents the same image but in different spectra. Note that in the work by Wang and Kumar [25], the authors considered that in the Cross-Eyed database, non-synchronously spectrum images were obtained (based on the numbers of intra- and inter-class comparisons), so they matched NIR against VIS images of the same index in the intra-class comparison. Then for a fair comparison with the state-of-the-art method [25], in the closed-world protocol, we also report results considering that the NIR and VIS images where obtained non-synchronously in the Cross-Eyed database.

In order to evaluate the robustness of the proposed methodology, we also evaluate and then report results on the *open-world* protocol, in which the training and test sets have images from different classes. In other words, there are no images from the same subject in the training and testing. In this protocol, for both databases, we use the first half of the subject images for training and the second half for testing.

The distributions of images and classes in the training and test sets, as well as the number of genuine and impostors pairs generated in the test phase for both databases and protocols are detailed in Table 1.

Table 1. Genuine and impostor matches for the Closed-world (CW) and Open-world (OW) protocols on Cross- and Intra-spectral scenarios. *The comparison with the state-of-the-art methods was performed using the closed-world protocol.

Database	Protocol	Scenario	Train/Test Images(Classes)	Gen./Imp. pairs
PolyU	CW	Cross	8,360(418)/4,180(418)	4,180/4,357,650
PolyU	CW	Intra	8,360(418)/2,090(418)	4,180/2,178,825
PolyU	OW	Cross	6,270(209)/6,270(209)	21,945/9,781,200
PolyU	OW	Intra	6,270(209)/3,135(209)	21,945/4,890,600
Cross-Eyed	CW	Cross	2,400(240)/1,440(240)	720/516,240
Cross-Eyed	CW	Intra	2,400(240)/720(240)	720/258,120
Cross-Eyed	OW	Cross	1,920(120)/1,920(120)	3,360/913,920
Cross-Eyed	OW	Intra	1,920(120)/960(120)	3,360/456,960

The mean and standard deviation of 30 repetitions for the EER and decidability figures obtained by the proposed methodology are shown.

5. RESULTS AND DISCUSSION

In this section, we present and discuss the results observed for the intra-spectral cross-spectral scenarios, in both the iris and periocular modalities. We start by providing the results using the closed-world protocol, in order to establish a baseline with respect to the state-of-the-art. We also investigate the impact of the feature vector size and the weights used to merge information from the periocular region and iris traits. Then, the results using the open-world protocol are presented, to perceive how robust deep representations can be obtained. Using the ResNet-50 model, a comparison of the verification effectiveness using features extracted from various network depths is performed. Lastly, we performed a subjective analysis of the pairwise errors.

In a complementary setting, we explore the advantages yielding from fusing representations of the periocular and iris traits to

improve performance. Similar to previous works [6, 43, 44] that applied higher weights in the most discriminating traits, and also considering that in all our experiments the periocular region reported better results compared to the iris, we decided to use constant weights of 0.6 and 0.4 respectively for the periocular and iris representations when obtaining the fused score by linear combination.

The experiments performed in this work and reported here used an NVIDIA® Titan Xp GPU with 12GB memory and 3,840 CUDA cores, and the tensorflow™ and *Keras* frameworks were used to implement the CNN models.

5.1. Closed-world protocol

At first, Table 2 and Table 3 report the results observed for verification mode, in the cross-spectral and intra-spectral scenarios (NIR against NIR and VIS against VIS) and using the closed-world protocol. In a way similar to Nalla and Kumar [6] and also to guarantee a fair comparison to their method, the fusion of two spectra on the PolyU database was carried out by linear combination, using weights of 0.6 and 0.4, respectively, to the NIR and VIS images. However, based on the individual spectral results, on the Cross-Eyed database, we used weights of 0.6 and 0.4 for the VIS and NIR representations, respectively. Also, on the Cross-Eyed database, we can perceive that the spectral fusion using iris representations extracted by the VGG16 model reported lower results than using the only VIS spectral information. The results show that the representations obtained from NIR images presented a high EER value, which penalized the fusion of spectra. Therefore, lower weight for NIR representations may improve the fusion result. The results of those fusions are shown in Table 2 and Table 3 (VIS and NIR Fusion section).

Anyway, it can be seen that - for both databases - the proposed approach achieves better results than the state-of-the-art methods, both in the cross-spectral and in the intra-spectral scenarios even that the protocol used in this paper is more challenging. For example, in the PolyU database, we used images from all 209 subjects in the experiments, while the approaches proposed by Wang and Kumar [25], and Nalla and Kumar [6] used images from only 140 subjects. In the Cross-Eyed database, based on the number of pairs of intra-class comparisons reported in the experiments by Wang and Kumar [25], the authors considered that the database has images obtained non-synchronously. Images from the Cross-Eyed database were obtained using a dual sensor with a beam splitter, so the NIR and VIS images are acquired simultaneously. However, we visually verified that the images of the same index, i.e., those that should be the same one in the NIR and VIS, have a random shift in each spectrum. Thus, for a fair comparison with the state-of-the-art approaches, we report the results using both protocols, considering the images obtained synchronously and non-synchronously. Note that we collected the state-of-the-art results from the original papers [6, 25], i.e., we did not have implemented any approach from these works.

In terms of the CNN architectures, the ResNet-50 model reported lower EER values compared to the VGG16 model in all cases. However, in some cases, specifically in the PolyU database, the representations extracted with the VGG16 model obtained a better separation of intra- and inter-class distributions, as can be seen in their Decidability index.

The results show that in Cross-Eyed, the periocular modality achieves better results than the iris one. However, in the PolyU database, there is no significant difference between iris and periocular representations, mainly in the intra-spectral experiments. From a visual inspection analysis of the pairwise comparison errors (some examples are shown in Section 5.5), we perceive that in the PolyU

Table 2. Results - closed-world protocol on the PolyU database. *Using only 140 subjects from a total of 209.

Approach	Modality	EER (%)	Decidability
Cross-Spectral			
CNN with SDH [25]*	Iris	5.39	2.13
CNN with SDH [25]	Iris	12.41	–
VGG16 with SDH [25]*	Iris	4.85	–
Proposed VGG16	Iris	2.16 ± 0.16	5.23 ± 0.08
ResNet50 with SDH [25]*	Iris	7.17	–
Proposed ResNet50	**Iris**	$\mathbf{1.13 \pm 0.14}$	$\mathbf{5.17 \pm 0.08}$
Proposed VGG16	Periocular	1.80 ± 0.21	6.03 ± 0.20
Proposed ResNet50	**Periocular**	$\mathbf{0.78 \pm 0.09}$	$\mathbf{5.97 \pm 0.08}$
Proposed VGG16	Fusion	0.93 ± 0.10	6.97 ± 0.13
Proposed ResNet50	**Fusion**	$\mathbf{0.49 \pm 0.06}$	$\mathbf{6.75 \pm 0.08}$
VIS vs VIS			
Nalla and Kumar [6]*	Iris	6.56	–
Proposed VGG16	Iris	1.53 ± 0.12	6.27 ± 0.08
Proposed ResNet50	**Iris**	$\mathbf{0.78 \pm 0.08}$	$\mathbf{5.91 \pm 0.07}$
Proposed VGG16	Periocular	1.50 ± 0.16	6.63 ± 0.21
Proposed ResNet50	**Periocular**	$\mathbf{0.61 \pm 0.11}$	$\mathbf{6.57 \pm 0.08}$
Proposed VGG16	Fusion	0.76 ± 0.10	7.73 ± 0.14
Proposed ResNet50	**Fusion**	$\mathbf{0.35 \pm 0.06}$	$\mathbf{7.44 \pm 0.10}$
NIR vs NIR			
Nalla and Kumar [6]*	Iris	3.97	–
Proposed VGG16	Iris	1.21 ± 0.13	6.61 ± 0.10
Proposed ResNet50	**Iris**	$\mathbf{0.68 \pm 0.07}$	$\mathbf{6.05 \pm 0.07}$
Proposed VGG16	Periocular	1.56 ± 0.19	6.58 ± 0.21
Proposed ResNet50	**Periocular**	$\mathbf{0.68 \pm 0.10}$	$\mathbf{6.59 \pm 0.07}$
Proposed VGG16	Fusion	0.70 ± 0.11	7.86 ± 0.17
Proposed ResNet50	**Fusion**	$\mathbf{0.40 \pm 0.06}$	$\mathbf{7.54 \pm 0.09}$
VIS and NIR Fusion			
Nalla and Kumar [6]*	Iris	2.86	–
Proposed VGG16	Iris	1.01 ± 0.09	6.81 ± 0.08
Proposed ResNet50	**Iris**	$\mathbf{0.59 \pm 0.08}$	$\mathbf{6.29 \pm 0.07}$
Proposed VGG16	Periocular	1.36 ± 0.15	6.79 ± 0.21
Proposed ResNet50	**Periocular**	$\mathbf{0.56 \pm 0.10}$	$\mathbf{6.82 \pm 0.08}$
Proposed VGG16	Fusion	0.63 ± 0.10	8.05 ± 0.16
Proposed ResNet50	**Fusion**	$\mathbf{0.35 \pm 0.05}$	$\mathbf{7.75 \pm 0.10}$

database, some uncontrolled conditions present in the images such as pose, eye gaze, and rotation may penalize the quality of the periocular representations. These conditions are more controlled in the cross-eyed images. Also, Cross-Eyed images are smaller than PolyU images, so the iris region is even smaller, and the periocular images are better centralized based on the iris region in the Cross-Eyed and not in the PolyU database. Nevertheless, Cross-Eyed images present a more significant difference in color and illumination among classes, which makes them more distinct and may explain the better results in VIS against VIS comparisons than NIR against NIR.

5.2. Feature size and fusion weights analyses

In this section, we analyze and discuss the impact of feature vector size and the weights used for the fusion of the iris and periocular region representations.

As state in Section 3, we choose the feature size of 256 based on the experiments and results reported by Luz et al [18]. Therefore, we also performed some experiments creating new models with different sizes in the last layer before the softmax one, i.e., the layer used to extract the features (representations). The results of the fusion of iris and periocular representations extracted with these models

Table 3. Results - closed-world protocol on the Cross-Eyed database. *same protocol used by Wang and Kumar [25].

Approach	Modality	EER (%)	Decidability
Cross-spectral			
CNN with SDH [25]	Iris	6.34	2.54
VGG16 with SDH [25]	Iris	3.13	–
Proposed VGG16*	Iris	5.58 ± 0.59	3.87 ± 0.16
Proposed VGG16	Iris	6.76 ± 0.56	3.58 ± 0.14
ResNet50 with SDH [25]	Iris	6.11	–
Proposed ResNet50*	**Iris**	**2.45 ± 0.25**	**4.73 ± 0.09**
Proposed ResNet50	**Iris**	**3.07 ± 0.38**	**4.49 ± 0.09**
Proposed VGG16*	Periocular	2.35 ± 0.28	5.61 ± 0.20
Proposed VGG16	Periocular	3.18 ± 0.42	5.19 ± 0.21
Proposed ResNet50*	**Periocular**	**1.45 ± 0.24**	**4.73 ± 0.09**
Proposed ResNet50	**Periocular**	**1.95 ± 0.35**	**5.34 ± 0.12**
Proposed VGG16*	Fusion	1.86 ± 0.19	5.78 ± 0.11
Proposed VGG16	Fusion	2.66 ± 0.29	5.31 ± 0.12
Proposed ResNet50*	**Fusion**	**1.06 ± 0.15**	**6.29 ± 0.11**
Proposed ResNet50	**Fusion**	**1.40 ± 0.26**	**5.93 ± 0.12**
VIS vs VIS			
Proposed VGG16	Iris	3.66 ± 0.39	4.85 ± 0.16
Proposed ResNet50	**Iris**	**2.47 ± 0.42**	**5.12 ± 0.13**
Proposed VGG16	Periocular	2.60 ± 0.40	5.57 ± 0.21
Proposed ResNet50	**Periocular**	**1.70 ± 0.37**	**5.66 ± 0.13**
Proposed VGG16	Fusion	1.94 ± 0.29	6.15 ± 0.16
Proposed ResNet50	**Fusion**	**1.17 ± 0.25**	**6.39 ± 0.13**
NIR vs NIR			
Proposed VGG16	Iris	7.31 ± 0.91	3.46 ± 0.18
Proposed ResNet50	**Iris**	**2.74 ± 0.34**	**4.72 ± 0.08**
Proposed VGG16	Periocular	2.97 ± 0.46	5.36 ± 0.23
Proposed ResNet50	**Periocular**	**1.78 ± 0.39**	**5.54 ± 0.13**
Proposed VGG16	Fusion	2.40 ± 0.35	5.36 ± 0.12
Proposed ResNet50	**Fusion**	**1.31 ± 0.24**	**6.14 ± 0.12**
VIS and NIR Fusion			
Proposed VGG16	Iris	3.69 ± 0.39	4.65 ± 0.15
Proposed ResNet50	**Iris**	**2.18 ± 0.31**	**5.25 ± 0.10**
Proposed VGG16	Periocular	2.44 ± 0.43	5.70 ± 0.22
Proposed ResNet50	**Periocular**	**1.54 ± 0.30**	**5.76 ± 0.13**
Proposed VGG16	Fusion	1.92 ± 0.29	6.09 ± 0.14
Proposed ResNet50	**Fusion**	**1.11 ± 0.20**	**6.47 ± 0.12**

Table 4. Feature vector size results fusing iris and periocular region traits on Cross-spectral scenario.

Model	Feat. Size	PolyU		Cross-Eyed	
		EER (%)	Decidability	EER (%)	Decidability
ResNet50	1024	0.54 ± 0.09	6.76 ± 0.10	1.61 ± 0.25	5.93 ± 0.13
	512	0.56 ± 0.06	6.73 ± 0.08	1.35 ± 0.22	6.00 ± 0.11
	256	0.49 ± 0.06	6.75 ± 0.08	1.40 ± 0.26	5.93 ± 0.12
	128	0.43 ± 0.05	6.70 ± 0.08	1.35 ± 0.30	5.99 ± 0.13
	64	0.37 ± 0.07	6.50 ± 0.08	1.26 ± 0.22	5.93 ± 0.15
	32	0.30 ± 0.05	6.05 ± 0.15	1.41 ± 0.27	5.65 ± 0.16
VGG16	1024	0.99 ± 0.10	6.85 ± 0.08	2.68 ± 0.28	5.29 ± 0.11
	512	0.92 ± 0.12	6.94 ± 0.11	2.53 ± 0.38	5.35 ± 0.14
	256	0.93 ± 0.10	6.97 ± 0.13	2.66 ± 0.29	5.31 ± 0.12
	128	0.80 ± 0.12	7.03 ± 0.10	2.78 ± 0.33	5.28 ± 0.10
	64	0.73 ± 0.11	6.93 ± 0.11	2.67 ± 0.37	5.23 ± 0.15
	32	0.69 ± 0.10	6.46 ± 0.07	2.79 ± 0.47	4.98 ± 0.17

and w_i stand for the periocular and iris weights, respectively. The results are reported in Figure 3.

Even though the values of EER are lower using features extracted with the ResNet50 model, we can observe a similar behaviour regarding the weight difference in both databases for both models. That is, when the weights are appropriately combined the best results are achieved. We can also observe that the periocular trait has more impact on the Cross-Eyed database than on the PolyU database. We also note that on the PolyU database, in some cases, fusion with a higher iris weight ($w_i = 0.6$ and $w_p = 0.4$ using VGG16 features) may achieve a lower value of EER.

5.3. Open-world protocol

Also, the experimental results observed for the open-world scenario are presented in Table 5 and Table 6 for the PolyU and Cross-Eyed databases, respectively. Notice that this protocol is more challenging since there is no sample of the test classes in the training set. Another factor that makes it more difficult is that compared to the closed-world protocol, fewer images are available for model training, and there are more images on the test set increasing the pair of genuine and imposter comparisons.

Table 5. Verification in the open-world protocol on the PolyU database.

Approach	Modality	EER (%)	Decidability
Cross-spectral			
Proposed ResNet50	Iris	12.01 ± 0.78	2.44 ± 0.08
Proposed ResNet50	Periocular	8.02 ± 0.65	3.00 ± 0.11
Proposed ResNet50	Fusion	6.01 ± 0.39	3.35 ± 0.08
VIS vs VIS			
Proposed ResNet50	Iris	4.30 ± 0.24	3.86 ± 0.07
Proposed ResNet50	Periocular	3.94 ± 0.27	4.14 ± 0.09
Proposed ResNet50	Fusion	2.61 ± 0.11	4.71 ± 0.06
NIR vs NIR			
Proposed ResNet50	Iris	4.00 ± 0.24	3.88 ± 0.08
Proposed ResNet50	Periocular	4.00 ± 0.26	4.10 ± 0.10
Proposed ResNet50	Fusion	2.55 ± 0.17	4.68 ± 0.10

are presented in Table 4. Luz et al. [18] stated that for the cosine distance metric, high dimensional vectors resulted in better performance. Conversely, our results show that representations extracted with the ResNet50 model achieve lower values of EER when the feature vector is smaller. The same occurs in the VGG16 model features in the PolyU database. Regarding the decidability index, the size of the feature vector does not show to have much impact. These results may be related to the fact that both models can generate sparse feature vectors, as stated by Wang and Kumar [25]. Thus a bigger feature vector will not always improve the performance of the biometric system. Here, we decided to keep a feature vector size of 256 because it keep a trade-off between EER and Decidability.

As described in Section 3, similar to some approaches [6, 43, 44] in the literature and based on the individual performance in our experiments, we choose weights of 0.6 and 0.4 for the periocular and iris fusion, respectively. Nevertheless, in this section, we evaluated the impact of different iris and periocular weights on the trait representations fusion in the cross-spectral scenario, for both models. Indeed, we impose $w_p \in [0, 1]$, such that $w_i + w_p = 1$, where w_p
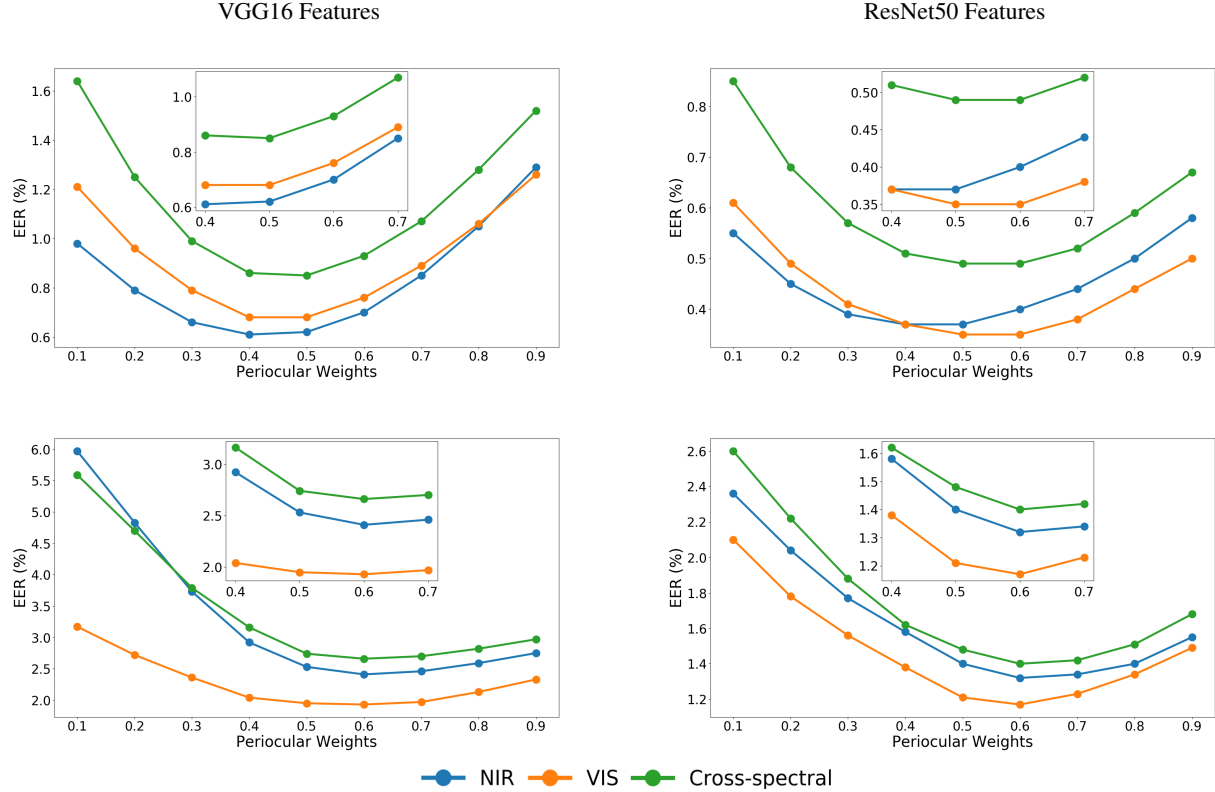
Fig. 3. Periocular weights impact on the traits fusion in the cross-spectral scenario on the PolyU (top row) and Cross-Eyed (bottom row) databases.

Table 6. Results - open-world protocol on the Cross-Eyed database.

Approach	Modality	EER (%)	Decidability
Cross-spectral			
Proposed ResNet50	Iris	8.87 ± 0.77	2.85 ± 0.11
Proposed ResNet50	Periocular	4.39 ± 0.44	3.85 ± 0.11
Proposed ResNet50	Fusion	3.51 ± 0.32	4.17 ± 0.07
VIS vs VIS			
Proposed ResNet50	Iris	4.25 ± 0.35	4.01 ± 0.10
Proposed ResNet50	Periocular	3.41 ± 0.38	4.41 ± 0.11
Proposed ResNet50	Fusion	2.57 ± 0.26	4.97 ± 0.09
NIR vs NIR			
Proposed ResNet50	Iris	5.04 ± 0.43	3.63 ± 0.12
Proposed ResNet50	Periocular	3.51 ± 0.40	4.38 ± 0.12
Proposed ResNet50	Fusion	2.75 ± 0.28	4.83 ± 0.10

To perceive the differences in performance, a comparison of the results using closed- and open-world is shown with the Receiver Operating Characteristic (ROC) curve in Figure 4. Even though a fully fair comparison between closed- and open-world protocols is not feasible because the number of subjects used for learning is different, it is noticeable that the open-world protocol reported worse performance in all modes compared to the closed-world protocol. Nevertheless, we conclude that fusing the ocular and iris represen-

tations also leads to promising results in the open-world protocol, given that the observed decidability was higher than three for both databases considered.

5.4. ResNet-50: Performance vs. Network Depth

Having concluded that the *ResNet-50* yields to the optimal results in terms of EER in our experiments, our next goal was to perceive how the verification performance varies with respect to the depth of the layer from where representations are taken. In this experiment, we considered all the convolution layers with stride equal to 2, resulting in four different depths to be tested: 12, 24, 42 and 50 layers. For each one of the four possibilities (depths), the same modifications described in the methodology section were made, adding a fully-connected layer with 256 neurons and a layer with a softmax cross-entropy loss function. The verification results using the different depths are reported in Table 7 for the PolyU and Cross-Eyed databases.

It can be observed that the largest degradation of the results occurred when using shallow models occurs in the Cross-Eyed database. In all cases, the VIS against VIS comparison reports the best results and it is the scenario where it presents the lowest degradation of the response in the different depths of the model.

As shown in the NIR against NIR and Cross-spectral results in the Cross-Eyed database, some EER values in the fusion of traits is higher than the ones using information from the periocular region only. This behavior is due to the weight used in the fusion of features where the low discrimination of the iris region penalizes and degrades the fused matching score, as we discuss in Section 5.2.
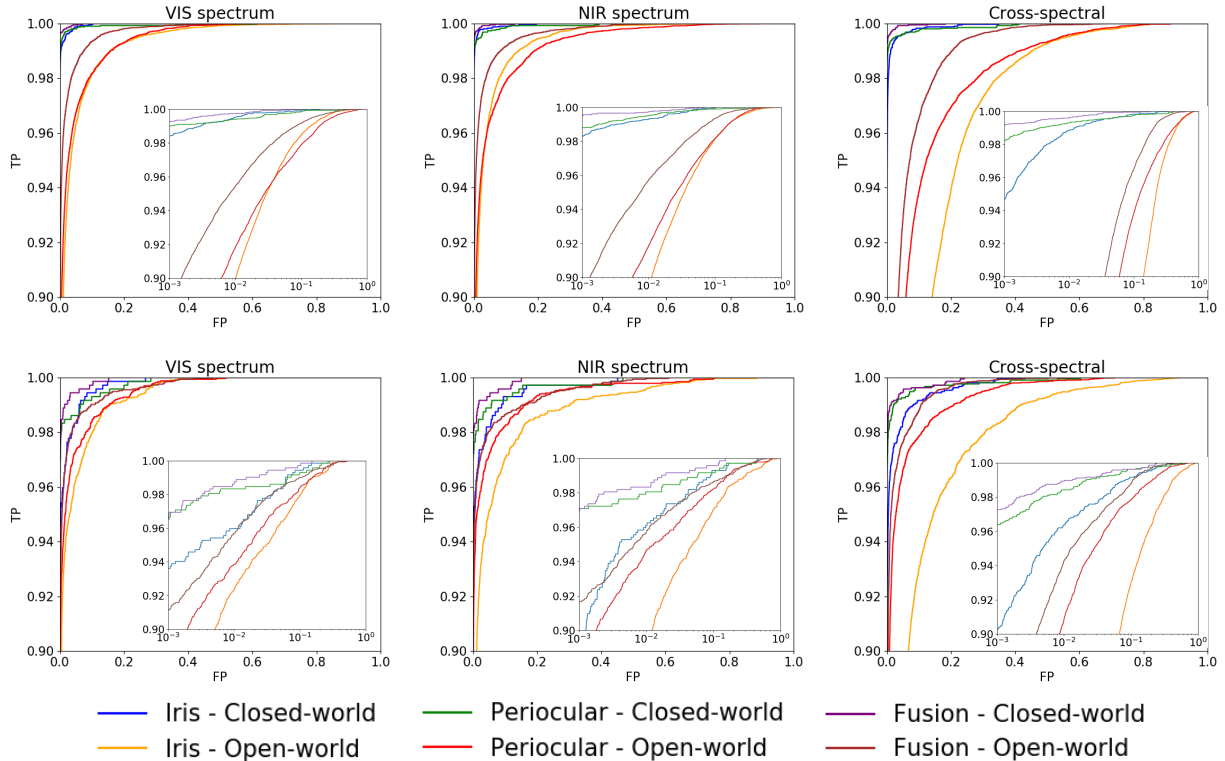
Fig. 4. ROC curves comparing the closed- and open-world protocols on the PolyU (top row) and Cross-Eyed (bottom row) databases.

Table 7. EER values observed for different depths (trainable parameters) of ResNet50 architecture, using the closed-world protocol.

Spec.	Trait	12 layers (26M)	24 layers (14.5M)	42 layers (15.6M)	50 layers (24.1M)
		PolyU			
VIS	Iris	3.21 ± 0.16	2.29 ± 0.15	1.60 ± 0.10	0.78 ± 0.08
VIS	Perioc.	3.84 ± 0.14	3.17 ± 0.18	2.17 ± 0.12	0.61 ± 0.11
VIS	Fusion	1.66 ± 0.06	1.41 ± 0.07	1.06 ± 0.11	0.35 ± 0.06
NIR	Iris	3.55 ± 0.18	2.36 ± 0.11	1.46 ± 0.10	0.68 ± 0.07
NIR	Perioc.	4.16 ± 0.17	3.39 ± 0.18	2.27 ± 0.14	0.68 ± 0.10
NIR	Fusion	2.13 ± 0.14	1.56 ± 0.08	1.09 ± 0.10	0.40 ± 0.06
Cross	Iris	6.39 ± 0.41	4.50 ± 0.23	3.09 ± 0.19	1.13 ± 0.14
Cross	Perioc.	5.38 ± 0.20	4.04 ± 0.17	2.71 ± 0.14	0.78 ± 0.09
Cross	Fusion	2.95 ± 0.15	2.07 ± 0.13	1.41 ± 0.09	0.49 ± 0.06
		Cross-Eyed			
VIS	Iris	4.77 ± 0.38	3.29 ± 0.26	2.16 ± 0.34	2.47 ± 0.42
VIS	Perioc.	6.34 ± 0.36	3.70 ± 0.35	1.90 ± 0.23	1.70 ± 0.37
VIS	Fusion	3.78 ± 0.22	1.94 ± 0.16	1.25 ± 0.18	1.17 ± 0.25
NIR	Iris	20.24 ± 0.70	16.28 ± 0.66	8.78 ± 0.56	2.74 ± 0.34
NIR	Perioc.	7.28 ± 0.35	4.08 ± 0.32	1.88 ± 0.23	1.78 ± 0.39
NIR	Fusion	7.78 ± 0.30	4.90 ± 0.33	2.03 ± 0.23	1.31 ± 0.24
Cross	Iris	20.88 ± 0.74	15.91 ± 0.60	8.12 ± 0.63	3.07 ± 0.38
Cross	Perioc.	7.53 ± 0.38	4.17 ± 0.38	2.31 ± 0.31	1.95 ± 0.35
Cross	Fusion	8.29 ± 0.46	4.43 ± 0.29	2.14 ± 0.24	1.40 ± 0.26

The experiments performed by Nguyen et al. [23] show that features extracted from intermediate layers of the networks achieved better results compared to deep layer representations. However, our results report lower EER rates using features extracted from deeper layers. It is important to point out that in [23] the ResNet152 model (i.e., a deeper model than ResNet50, used in our work) was employed. The same behavior can be observed in work by Henandez-Diaz et al. [31], were the authors stated that features extracted from the intermediate layers of the ResNet-101 model reported the best results. Thus, the deepest layer reported in this work is approximately at the same depth as the intermediate layer reported by Nguyen et al. [23] and by Hernandez-Diaz et al. [31]. In another work, Hernandez-Diaz et al. [46] reported that using the ResNet50 model, representations from the intermediate layers achieved better results in the UBIPr Periocular database [47]. Oppositely, in this work, periocular representations extracted from the last layer of the ResNet50 model achieved the best results. Notice the UBIPr database has some larger images (from 501×401 pixels (8m) to 1001×801 (4m)) than PolyU and Cross-Eyed databases and also the periocular region is more extensive, containing eyebrows information, which can explain why a shallow model can extract more discriminant features from the intermediate layers, in this case.

As described in [25], a disadvantage of the VGG16 model, when compared to ResNet, is its larger number of trainable parameters (98.6M, when compared to their CNN with SD methodology 0.6M). As before stated, in our case the best responses were observed when using the ResNet50 model, which after the modifications has 24.1M (four times lower compared to VGG16). As shown in Table 7, smaller networks in terms of depth lead to increasingly high losses in performance, however also decreasing nearly 10M training parameters, which can be an interesting solution for embedded systems and

other cases where the computational complexity might be a concern. The ResNet with 12 layers has more trainable parameters than the other models, since it considers an input image of 28×28 pixels and 128 filters. In addition, its convolutional part is connected with a fully connected layer containing 256 neurons added for reduction of feature dimensionality.

5.5. Subjective evaluation

In order to provide some insight about the weaknesses of the solutions proposed in this paper, and also to provide a basis for subsequent improvements in the technology, this section highlights some notable cases of image pairwise comparisons that led to the best/worst performance (using the closed-world protocol). Results are shown in Figure 5, grouped into the worst genuine (when the system rejected a true matching) and the best impostors (when the system accepted a false matching) comparisons.
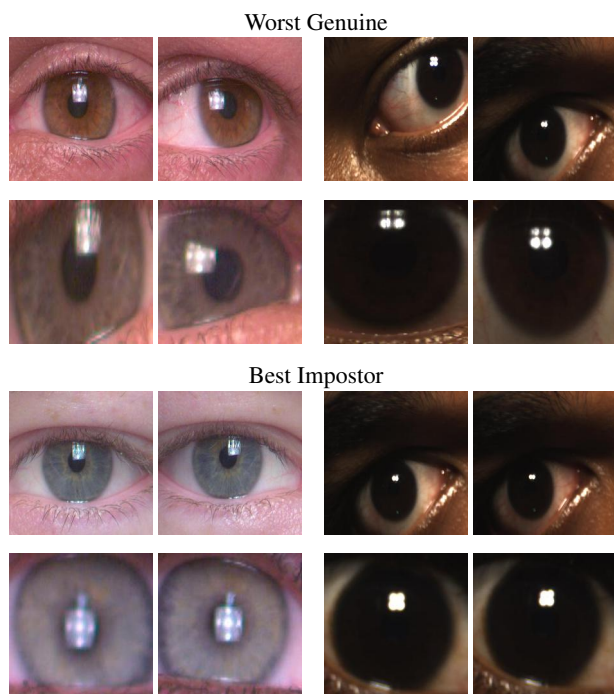
Fig. 5. Pairwise comparison errors in the VIS against VIS scenario on Cross-Eyed (left) and PolyU (right) databases. Periocular and iris matching modalities are presented at Top and Bottom rows, respectively.

Although Figure 5 only shows VIS images, we noticed that pose and gaze are factors that can lead to matching errors also in NIR against NIR and cross-spectral scenarios. We observe that there were also confusions in images of the same subject but from different classes (left and right eyes) no matter the spectral scenario. Thus, we believe that it is possible to improve the recognition system accuracy using information based on the angle of the periocular region images and also performing a preprocessing to determine the left and right eyes (i.e., a soft biometrics process). Also, based on the pairwise comparison errors, we can state that another factor that may improve system accuracy is the process of centralization/resizing of the periocular image based on the iris region size and location, similar to the method proposed by Hernande-Diaz et al. [46].

6. CONCLUSION

In this work we performed extensive experiments on two databases for both cross-spectral and intra-spectral ocular recognition. A strategy using methodologies from the literature was applied to reach new state-of-the-art results on both databases. It shows that there is still room for improvement by applying and merging known methodologies in the literature to surpass cross-spectral ocular recognition.

We also discuss how deep representations from the iris and ocular region (extracted using VGG16 and ResNet50 architectures) can be fused to improve the recognition performance on the ambitious cross-spectral recognition problem. We used CNN models that were pre-trained for face recognition, and fine-tuned each one for a specific biometric modality: iris and periocular. A single model for each trait was trained for the feature extraction using NIR and VIS images. The matching phase, on a verification mode, was performed using the cosine metric. In order to provide a fair comparison with the state-of-the-art approaches, we used the closed-world protocol. However, we also reported results on the open-world protocol to evaluate the robustness of the proposed methodology.

Our experiments showed that the models learned on the ResNet-50 architecture reported best results in terms of EER than its VGG counterpart, both in the PolyU and Cross-Eyed databases. Interestingly, we note that even this simple processing chain was observed to advance the state-of-the-art results in both datasets.

Overall, in most of the experiments, features taken from the periocular region were observed to provide better performance than iris features, with the fusion of these two modalities improving the EER value and decidability index than the best individual trait.

In a complementary way, we analyzed the impact of the feature vector size and the Iris and Periocular weights used for trait representation fusion, and how the recognition performance varies with respect to the depth of the models used for feature extraction, i.e., by using intermediate layers of the ResNet50 model to take the feature sets used in the matching phase.

Finally, subjective analysis of the best/worst false genuine and true impostors image pairwise comparisons was also performed, showing that factors such as angles of image capture may interfere in the accuracy of the recognition system. In this direction, we plan for future works to investigate how to build representation taking into account eye gaze and pose.

Acknowledgment

References

[1] H. Proença and L. A. Alexandre, "Toward covert iris biometric recognition: Experimental results from the NICE contests," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, 2012.

[2] H. Proença and L. A. Alexandre, "UBIRIS: A noisy iris image database," in *13th International Conference on Image Analysis and Processing - ICIAP 2005*, 2005, vol. 3617, pp. 970–977.

[3] H. Proença, S. Filipe, R. Santos, J. Oliveira, and L. A. Alexandre, "The UBIRIS.v2: A database of visible wavelength iris images captured on-the-move and at-a-distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1529–1535, 2010.

[4] M. S. Hosseini, B. N. Araabi, and H. Soltanian-Zadeh, "Pigment melanin: Pattern for iris recognition," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 4, pp. 792–804, 2010.

[5] A. Sharma, S. Verma, M. Vatsa, and R. Singh, "On cross spectral periocular recognition," *2014 IEEE International Conference on Image Processing, ICIP 2014*, pp. 5007–5011, 2014.

[6] P. R. Nalla and A. Kumar, "Toward more accurate iris recognition using cross-spectral matching," *IEEE Transactions on Image Processing*, vol. 26, no. 1, pp. 208–221, 2017.

[7] F. M. Algashaam, K. Nguyen, M. Alkanhal, V. Chandran, W. Boles, and J. Banks, "Multispectral Periocular Classification With Multimodal Compact Multi-Linear Pooling," *IEEE Access*, vol. 5, pp. 14572–14578, 2017.

[8] G. Hinton, L. Deng, D. Yu, G. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, and B. Kingsbury, "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.

[9] Y. Zhang, W. Chan, and N. Jaitly, "Very deep convolutional networks for end-to-end speech recognition," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2017, pp. 4845–4849, IEEE.

[10] S. Kim, T. Hori, and S. Watanabe, "Joint CTC-attention based end-to-end speech recognition using multi-task learning," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2017, pp. 4835–4839, IEEE.

[11] X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach," in *Proceedings of the 28th International Conference on International Conference on Machine Learning*, 2011, pp. 513–520.

[12] R. Socher, J. Pennington, E. H. Huang, A. Y. Ng, and C. D. Manning, "Semi-supervised recursive autoencoders for predicting sentiment distributions," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 2011, pp. 151–161, Association for Computational Linguistics.

[13] R. Laroca, E. Severo, L. A. Zanlorensi, L. S. Oliveira, G. R. Gonçalves, W. R. Schwartz, and D. Menotti, "A robust real-time automatic license plate recognition based on the yolo detector," in *2018 International Joint Conference on Neural Networks (IJCNN)*, July 2018, pp. 1–10.

[14] A.G. Hochuli, L.S. Oliveira, A.S. Britto Jr, and R. Sabourin, "Handwritten digit segmentation: Is it still necessary?," *Pattern Recognition*, vol. 78, pp. 1 – 11, 2018.

[15] R. Laroca, L. A. Zanlorensi, G. R. Gonçalves, E. Todt, W. R. Schwartz, and D. Menotti, "An efficient and layout-independent automatic license plate recognition system based on the YOLO detector," *arXiv preprint*, vol. arXiv:1909.01754, pp. 1–14, 2019.

[16] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *British Machine Vision Conference (BMVC)*, 2015, pp. 1–12.

[17] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," *CoRR*, 2017.

[18] E. Luz, G. M., L. A. Zanlorensi Junior, and D. Menotti, "Deep periocular representation aiming video surveillance," *Pattern Recognition Letters*, vol. 114, pp. 2 – 12, 2018.

[19] H. Proença and J. C. Neves, "Deep-PRWIS: Periocular Recognition Without the Iris and Sclera Using Deep Learning Frameworks," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 4, pp. 888–896, 2018.

[20] 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, pp. 154–161, 2016.

[21] A. Gangwar and A. Joshi, "DeepIrisNet: Deep iris representation with applications in iris recognition and cross-sensor iris recognition," in *IEEE Intern. Conference on Image Processing*, 2016, pp. 2301–2305.

[22] A. S. Al-Waisy, R. Qahwaji, S. Ipson, S. Al-Fahdawi, and T. A. M. Nagem, "A multi-biometric iris recognition system based on a deep learning approach," *Pattern Analysis and Applications*, 2017.

[23] K. Nguyen, C. Fookes, A. Ross, and S. Sridharan, "Iris recognition with off-the-shelf CNN features: A deep learning perspective," *IEEE Access*, vol. 6, pp. 18848–18855, 2018.

[24] H. Proença and J. C. Neves, "IRINA: Iris Recognition (Even) in Inaccurately Segmented Data," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, vol. 2017-Janua, pp. 6747–6756.

[25] K. Wang and A. Kumar, "Cross-spectral iris recognition using cnn and supervised discrete hashing," *Pattern Recognition*, vol. 86, pp. 85 – 98, 2019.

[26] L. A. Zanlorensi, E. Luz, R. Laroca, A. S. Britto Jr., L. S. Oliveira, and D. Menotti, "The impact of preprocessing on deep representations for iris recognition on unconstrained environments," in *Conference on Graphics, Patterns and Images (SIBGRAPI)*, Oct 2018, pp. 289–296.

[27] M. De Marsico, A. Petrosino, and S. Ricciardi, "Iris recognition through machine learning techniques: A survey," *Pattern Recognition Letters*, vol. 82, pp. 106 – 115, 2016, An insight on eye biometrics.

[28] J. Deng, W. Dong, R. Socher, L. J. Li, Kai Li, and Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[29] J. Daugman, "How iris recognition works," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 21–30, 2004.

[30] P. H. Silva, E. Luz, L. A. Zanlorensi, D. Menotti, and G. Moreira, "Multimodal feature level fusion based on particle swarm optimization with deep transfer learning," in *2018 Congress on Evolutionary Computation (CEC)*, 2018, pp. 1–8.

[31] F. Alonso-Fernandez K. Hernandez-Diaz and J. Bigun, "Cross spectral periocular matching using resnet features," in *International Conference on Biometrics(ICB)*, 2019, pp. 1–6, In Press.

[32] A. Sequeira, L. Chen, P. Wild, J. Ferryman, F. Alonso-Fernandez, K. B. Raja, R. Raghavendra, C. Busch, and J. Bigun, "Cross-Eyed - Cross-Spectral Iris/Periocular Recognition Database and Competition," in *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2016, vol. P-260, pp. 1–5.

[33] A. F. Sequeira, L. Chen, J. Ferryman, P. Wild, F. Alonso-Fernandez, J. Bigun, K. B. Raja, R. Raghavendra, C. Busch, T. de Freitas Pereira, S. Marcel, S. S. Behera, M. Gour, and V. Kanhangad, "Cross-eyed 2017: Cross-spectral iris/periocular recognition competition," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*, Oct 2017, pp. 725–732.

[34] D. Menotti, G. Chiachia, A. Pinto, W. R. Schwartz, H. Pedrini, A. X. Falcão, and A. Rocha, "Deep representations for iris, face, and fingerprint spoofing detection," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 864–879, 2015.

[35] Y. Du, T. Bourlai, and J. Dawson, "Automated classification of mislabeled near-infrared left and right iris images using convolutional neural networks," in *IEEE International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 2016, pp. 1–6.

[36] L. He, G. Poggi, C. Sansone, L. Verdoliva, H. Li, F. Liu, N. Liu, Z. Sun, and Z. He, "Multi-patch convolution neural network for iris liveness detection," in *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. 2016, pp. 1–7, IEEE.

[37] E. Severo, R. Laroca, C. S. Bezerra, L. A. Zanlorensi, D. Weingaertner, G. M., and D. Menotti, "A benchmark for iris location and a deep learning detector evaluation," in *2018 International Joint Conference on Neural Networks (IJCNN)*, July 2018, pp. 1–7.

[38] D. R. Lucio, R. Laroca, L. A. Zanlorensi, G. Moreira, and D. Menotti, "Simultaneous iris and periocular region detection using coarse annotations," in *Conference on Graphics, Patterns and Images (SIBGRAPI)*, Oct 2019, pp. 1–8, In Press.

[39] D. R. Lucio, R. Laroca, E. Severo, A. S. Britto Jr., and D. Menotti, "Fully convolutional networks and generative adversarial networks applied to sclera segmentation," in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, Oct 2018, pp. 1–7.

[40] C. S. Bezerra, R. Laroca, D. R. Lucio, E. Severo, L. F. Oliveira, A. S. Britto Jr., and D. Menotti, "Robust iris segmentation based on fully convolutional networks and generative adversarial networks," in *Conference on Graphics, Patterns and Images (SIBGRAPI)*, Oct 2018, pp. 281–288.

[41] J. Tapia and C. Aravena, "Gender classification from nir iris images using deep learning," in *Deep Learning for Biometrics*, 2017, pp. 219–239.

[42] Francesco M. et al., "A deep learning approach for iris sensor model identification," *Pattern Recognition Letters*, 2017.

[43] N. U. Ahmed, S. Cvetkovic, E. H. Siddiqi, A. Nikiforov, and I. Nikiforov, "Using fusion of iris code and periocular biometric for matching visible spectrum iris images captured by smart phone cameras," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, Dec 2016, pp. 176–180.

[44] N. U. Ahmed, S. Cvetkovic, E. H. Siddiqi, A. Nikiforov, and I. Nikiforov, "Combining iris and periocular biometric for matching visible spectrum eye images," *Pattern Recognition Letters*, vol. 91, pp. 11 – 16, 2017, Mobile Iris CHallenge Evaluation (MICHE-II).

[45] J. Daugman, "The importance of being random: statistical principles of iris recognition," *Pattern Recognition*, vol. 36, no. 2, pp. 279–291, 2003.

[46] K. Hernandez-Diaz, F. Alonso-Fernandez, and J. Bigun, "Periocular recognition using cnn features off-the-shelf," in *2018 International Conference of the Biometrics Special Interest Group (BIOSIG)*, Sep. 2018, pp. 1–5.

[47] C. N. Padole and H. Proenca, "Periocular recognition: Analysis of performance degradation factors," in *IAPR International Conference on Biometrics (ICB)*, New Delhi, India, mar 2012, pp. 439–445, IEEE.