Segmentation-less and Non-holistic Deep-Learning Frameworks for Iris Recognition

Hugo Proença and João C. Neves
IT - Instituto de Telecomunicações
University of Beira Interior, Portugal
{hugomcp,jcneves}@di.ubi.pt

Abstract

*Driven by the pioneer iris biometrics approach, the most relevant recognition methods published over the years are "*phase-based*", and segment/normalize the iris to obtain dimensionless representations of the data that attenuate the differences in scale, translation, rotation and pupillary dilation. In this paper we present a recognition method that dispenses the iris segmentation, noise detection and normalization phases, and is agnostic to the levels of pupillary dilation, while maintaining state-of-the-art performance. Based on deep-learning classification models, we analyze the displacements between* biologically corresponding *patches in pairs of iris images, to discriminate between* genuine *and* impostor *comparisons. Such corresponding patches are firstly learned in the normalized representations of the irises - the domain where they are optimally distinguishable - but are remapped into a* segmentation-less *polar coordinate system that uniquely requires iris detection. In recognition time, samples are only converted into this* segmentation-less *coordinate system, where matching is performed. In the experiments, we considered the challenging* open-world *setting, and used three well known data sets (CASIA-4-Lamp, CASIA-4-Thousand and WVU), concluding positively about the effectiveness of the proposed algorithm, particularly in cases where accurately segmenting the iris is a challenge.*

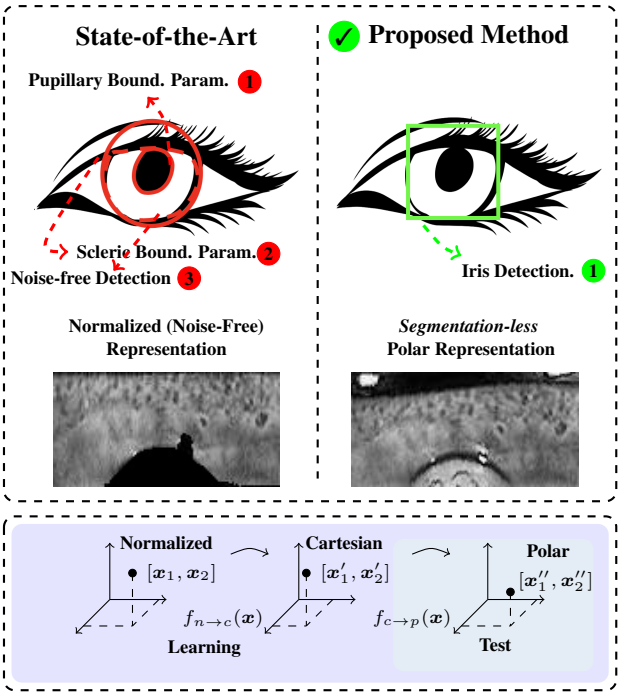


Figure 1. We propose a non-holistic iris recognition method that doesn't require iris segmentation and is agnostic to the levels of pupillary dilation. During learning time, we infer the appearance of *biologically corresponding* patches in *segmentation-less* representations of the iris data, which are the uniques used in recognition (test) time.

1. Introduction

After almost three decades of research [4], iris recognition under controlled conditions tends to be considered a solved problem, with most efforts being now concentrated in augmenting the recognition robustness [2].

Deep-learning frameworks were unquestionably a breakthrough in many computer vision tasks, from image segmentation [16], to object detection [32] and classification [14]. Being data-driven models, these frameworks do not depend on human efforts to specify image features, upon the availability of large amounts of data. In the biometrics context, many deep-learning based recognition methods have been proposed, based in the face [6] [10], iris/ocular [39] [26] or finger vein [27] traits. Despite the remarkable ability of this kind of frameworks to model complex data patterns, demanding them to autonomously infer the concept of *identity* might be too ambitious, particularly when working with poor-quality data.

This paper describes one iris recognition algorithm that - singularly - uses deep learning frameworks exclusively to perceive what are *biologically corresponding* patches in pairs of images, which is an evidently *easier* task than inferring the identity at once. In this sense, our work can be regarded as an evolution of the concept behind IRINA [25], with one major novelty: in test (recognition) time, we don't even segment the pupillary and scleric iris boundaries nor discriminate the occluded parts of the iris. Instead, we rely uniquely in one iris detection step to obtain a *segmentation-less* representation of the iris data, where all subsequent analysis is carried out.

Note that previous works have also proposed *segmentation-less* algorithms, but they were specifically designed for different traits, such as the ocular region (e.g., [28] and [29]), where changes in signals' *phase* are not as problematic as in the case of the iris. Also, the existing methods concerned about the misalignments between iris patches (e.g., [34]) approached the problem from parametric perspectives, and working in the polar representations of the iris data.

1.1. Contributions

This paper we describe one iris recognition algorithm that offers three contributions:

- we describe a novel *segmentation-less* polar representation of the iris data. Even though this representation only partially attenuates the differences in pupillary dilation and scale, it provides novel types of information that are useful for biometric recognition (e.g., the shape of the pupil);
- we propose a processing chain that depends uniquely of a bounding box ($[\boldsymbol{x}, r] \in \mathbb{N}^3$) provided by an iris detector (as a preprocessing step), which is in opposition to previously published non-holistic recognition algorithms;
- the proposed algorithm is agnostic to the levels of pupillary dilation, which is also a novelty;

Finally, the use of deep-learning frameworks to infer *biologically corresponding* patches should also be considered a contribution. Note that the concept of *corresponding patch* lies in a hyperspace of much lower dimensionality than the space of *identities*, where deep learning frameworks typically operate. As such, the amount of labelled data required for appropriate learning of our model is kept relatively short.

1.2. Iris Recognition: Strides and Challenges

Strides in iris recognition have been concentrated in: 1) extending the data acquisition volume; 2) improving robustness to *unconstrained* conditions; 3) obtaining interpretable iris representations; 4) developing cancellable signatures; and 5) providing cross-sensor operability. In terms of the data acquisition volume, the *iris-on-the-move* system [19] can be highlighted, along with Hsieh *et al.*'s [11] work, based in wavefront coding and super-resolution techniques. Regarding the *unconstrained* recognition problem, Dong *et al.* [7] used adaptive matching to identify the best features per individual. Pillai *et al.* [23] sparsely represented randomly projected iris patches, while Yang *et al.* [36] relied in high-order information to match irises pairs, and Alonzo-Fernandez *et al.* [8] proposed a super-resolution method based on eigen-transformations of the iris patches. Bit consistency is a concern, with approaches selecting only parts of the biometric signatures for matching (e.g. [12], [33] and [17]). Recognition in mobile devices is a particular case of the *unconstrained recognition* problem, with two international evaluation contests being recently conducted: using visible wavelength (Marsico *et al.* [18]) and NIR data (Zhang *et al.* [40]). In terms of interpretability, Chen *et al.* [3] exclusively relied in features that are intuitively understood by humans. For cross-sensor recognition, Pillai et al. [24] learned kernel-based transformations between the data acquired by different devices. Finally, Zhao *et al.* [37] suggested the concept of negative recognition as a way to cancellable biometrics, using complementary information of the biometric data for matching.

Deep learning-based methods have been advancing the state-of-the-art, with even fine-tuning providing good results [21]. Solutions have been proposed either for particular phases of the recognition chain (e.g., segmentation [13] or spoofing detection [20]), or for the whole process: Gangwar and Joshi [9] claimed that very deep architectures are particularly efficient to model the microstructures of iris texture. Zhang *et al.* [41] developed a feature fusion strategy that considers the complementarity of the iris and periocular information. Zhao and Kumar [38] used fully convolutional networks to generate phase-based discriminative features, based on an extended triple loss that accounts for bit shifting and non-iris regions masking. Finally, Proença and Neves [25] analyzed the phase/magnitude of the free-form deformation fields (2D registration vectors) between images, using second-order statistics to discriminate between genuine and impostor pairs.

The remainder of this paper is organized as follows: Section 2 provides a detailed description of the proposed method. In Section 3 we discuss the obtained results and the conclusions are given in Section 4.

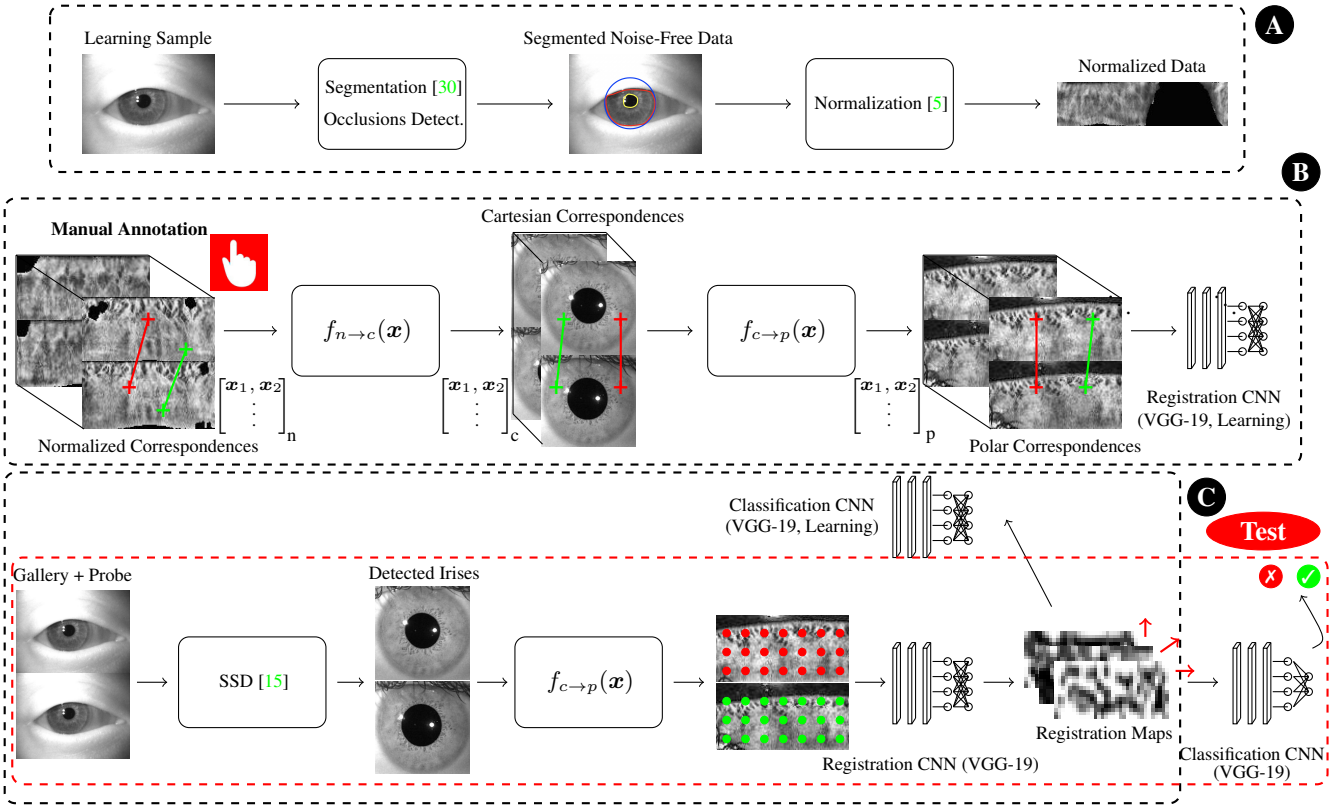


Figure 2. Cohesive perspective of the recognition method proposed in this paper: we accurately segment and normalize the noise-free iris regions of a learning set, and manually annotate the regions in *genuine* pairs that appear to regard the same biological region *corresponding iris patches*. Such correspondences are remapped to Cartesian coordinates and - next - to a pseudo-polar coordinate system that doesn't require image segmentation, where a CNN learns the notion of *corresponding* region.

2. Proposed Method

2.1. Learning Phase

We employ the term *biologically corresponding* iris patches to refer regions in pairs of *genuine* images that, under visual inspection, seem to regard the same biological region.

2.1.1 *Biologically Corresponding* Iris Patches

Let $\mathcal{I} = \{\boldsymbol{I}_1, \ldots, \boldsymbol{I}_n\}$ be a learning set of n iris images. For every $\boldsymbol{I}$, we obtain its shape-flexible and size-invariant normalized representation, as proposed by Daugman [5]. The mapping is given by $I\big(x(r,\theta), y(r,\theta)\big) \to I(r,\theta)$:

$$\begin{bmatrix} x(r,\theta) \\ y(r,\theta) \end{bmatrix} = \begin{bmatrix} x_p(\theta) & x_s(\theta) \\ y_p(\theta) & x_s(\theta) \end{bmatrix} \begin{bmatrix} 1-r \\ r \end{bmatrix}, \tag{1}$$

with the dimensionless parameter r spanning in the unit interval.

Next, for pairs of *genuine* normalized images $\{\boldsymbol{I}, \boldsymbol{I}'\}$, we manually annotate the control points that seem to regard the same iris feature. Groups of control points define two convex polygons $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}'$, represented by the coloured dots ($\boldsymbol{x}_i$ and $\boldsymbol{x}'_i$) in Fig. 3. Let $\boldsymbol{x}_i = (x_i, y_i)$ and $\boldsymbol{x}'_i = (x'_i, y'_i)$, $i = \{1, \ldots t\}$ be the locations of such pointwise correspondences. As in [25], we learn two functions f_1, f_2 that establish a dense set of correspondences between positions (rows/columns) in $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}'$, $f_1, f_2 : \mathbb{N}^2 \to \mathbb{N}$, such that $\forall\, \boldsymbol{x}'_i \in \boldsymbol{\Gamma}'$, $\boldsymbol{x}'_i = \big(f_c(\boldsymbol{x}), f_r(\boldsymbol{x})\big)$:

$$f_2(\boldsymbol{x}) = \boldsymbol{\lambda}_c^T [\boldsymbol{\phi}, p(\boldsymbol{x})], \tag{2}$$

$$f_1(\boldsymbol{x}) = \boldsymbol{\lambda}_r^T [\boldsymbol{\phi}, p(\boldsymbol{x})], \tag{3}$$

with $\boldsymbol{\phi} = \big[\phi\big(|\boldsymbol{x} - \boldsymbol{x}_1|_2\big), \ldots, \phi\big(|\boldsymbol{x} - \boldsymbol{x}_t|_2\big)\big]$, $|.|_2$ representing the ℓ_2 norm, $\phi(r) = e^{(-r/\kappa)^2}$ being a radial basis function and $p(\boldsymbol{x}) = [1, x, y]$ being a polynomial basis of first degree for a 2-dimensional vector space ($\kappa = 0.1$ was used in all our experiments). To obtain the $\boldsymbol{\lambda}$ coefficients, we define a $t \times t$ matrix $\boldsymbol{A}$, $A_{i,j} = \phi\big(|\boldsymbol{x}_i - \boldsymbol{x}_j|_2\big)$ and $\boldsymbol{P}$ as the $t \times 3$ polynomial basis matrix, such that $\boldsymbol{P} = [p(\boldsymbol{x}_1); \ldots; p(\boldsymbol{x}_t)]$. Then, $\boldsymbol{\lambda}_c$ and $\boldsymbol{\lambda}_r$ are given by:

$$\boldsymbol{\lambda}_c = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{P} \\ \boldsymbol{P}^T & \boldsymbol{0} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{x}'_c \\ \boldsymbol{0} \end{bmatrix} \quad (4)$$

$$\boldsymbol{\lambda}_r = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{P} \\ \boldsymbol{P}^T & \boldsymbol{0} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{x}'_r \\ \boldsymbol{0} \end{bmatrix} \quad (5)$$

with $\boldsymbol{x}'_c$=$[x'_1, \ldots, x'_t]^T$ and $\boldsymbol{x}'_r$=$[y'_1, \ldots, y'_t]^T$ concatenating the horizontal (column) and vertical (row) positions of the control points in $\boldsymbol{\Gamma}'$.

According to this procedure, we deem that any position $\boldsymbol{x} \in \boldsymbol{\Gamma}$ corresponds biologically to $\boldsymbol{x}' = \big(f_c(\boldsymbol{x}), f_r(\boldsymbol{x})\big) \in \boldsymbol{\Gamma}'$. As $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}'$ have different size and shape, this set of correspondences implicitly encodes the non-linear deformations that affect the iris texture, which is the core for the performance of the proposed method.

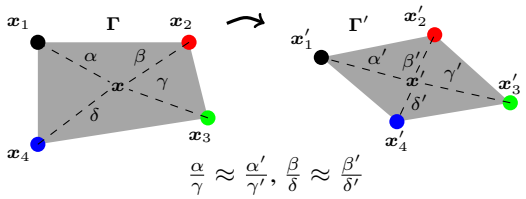


Figure 3. *Biologically corresponding* iris patches. Based on a set of manually annotated corresponding control points ($\boldsymbol{x}$, and $\boldsymbol{x}'$) in a pair of iris samples, two polygons ($\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}'$) are defined. For every position inside $\boldsymbol{\Gamma}$, the corresponding position in $\boldsymbol{\Gamma}'$ is found, with both points deemed to refer to the same biological region.

2.1.2 *Segmentation-less* Polar Representation

Next, we convert all positions in the normalized iris data into a segmentation-less polar space. Let $\boldsymbol{x} = (x, y)$ be a point in the $h \times w$ normalized representation of the iris. As an intermediate step, we map $\boldsymbol{x}$ into the original Cartesian space and from there into the segmentation-less space. Function $f_{n\to c}(.) : \mathbb{N}^2 \to \mathbb{N}^2$, performs the first mapping:

$$f_{n\to c}(\boldsymbol{x}) = \rho\, \boldsymbol{x}_p(\theta) + (1-\rho)\, \boldsymbol{x}_s(\theta), \quad (6)$$

with $\boldsymbol{x}_p$ and $\boldsymbol{x}_s$ denoting the coordinates of the pupillary and scleric borders at angle θ, and (θ, ρ) given by:

$$\theta = \frac{2\pi(x-1)}{w}, \rho = \frac{h-y}{h-1}. \quad (7)$$

Next, using $f_{c\to p}(\boldsymbol{x}) : \mathbb{N}^2 \to \mathbb{N}^2$, we map the position in the Cartesian space into the polar representation:

$$f_{c\to p}(\boldsymbol{x}) = \Big[\text{atan2}\left(\frac{\frac{s}{2}-y}{x-\frac{s}{2}}\right) w(2\pi)^{-1} + 1,$$
$$(|\boldsymbol{x} - [\frac{s}{2}, \frac{s}{2}]|_2 - \tau_p) h (x_c - \tau_p)^{-1}\Big], \quad (8)$$

with r_p denoting an average radius for the pupils and $s \times s$ is the size of the bounding box returned by an iris detection step. In this sense, $f_{c\to p}(f_{n\to c}(\boldsymbol{x}))$ converts points in the normalized iris data into the *segmentation-less* polar representation.

Finally, we consider patches $\boldsymbol{P}$ (from $\boldsymbol{\Gamma}$) and $\boldsymbol{P}'$ (from $\boldsymbol{\Gamma}'$) of 15×15 pixels, centered at each point correspondence. This set of patches is concatenated into a set of negative instances (non-corresponding patches), feeding the learning process of a Convolutional Neural Network ("Registration CNN"), responsible to discriminate between patches that correspond or not (part "B" in Fig. 2). The architecture of this CNN is based in the VGG-19 model, adapting the size of the input layer to depth 2, and removing all the pooling layers, due to the relatively small dimensions of the input data. The output layer is a *soft max* loss corresponding to the probability of two iris patches *correspond*. Learning was done according to the stochastic gradient descend algorithm, with an initial learning rate of 1e^{-2}, momentum set to 0.9 and weight decay equals to $1e^{-3}$.

The steps of the learning phase are shown in Fig. 2 as the "C" group. We use the popular SSD [13] detector in the learning data, to detect the irises in both images. The detector was adapted so to use default boxes with a single ratio, corresponding to squares. This way, the detector returns a parameterization $[\boldsymbol{x}_c, s] \in \mathbb{N}^3$, with $\boldsymbol{x}_c$ being the center of the $s \times s$ detected region.

Using $f_{c\to p}(\boldsymbol{x}), \forall \boldsymbol{x}$ we get the *segmentation-less* polar representation of the iris, where a $\tau_r \times \tau_c$ grid of equally spaced points is overlapped. Let $\boldsymbol{I}, \boldsymbol{I'}$ be a pair of irises in the *segmentation-less* polar space. Let $\boldsymbol{x}$ be the position of one grid point in the first image. We define a 15×15 patch around this point and densely sample the $\boldsymbol{x} \pm [\delta_r, \delta_c]$ positions ($\delta_r \in \{1, \ldots, \Delta_r\}, \delta_c \in \{1, \ldots, \Delta_c\}$) in the other image, and take 15×15 patches centered at each position. Pairs of patches are provided to the previously learned Registration CNN, with output scores directly corresponding to the probability that pairs of patches in the first and second image correspond biologically. By rearranging the CNN scores, we create the so-called *registration maps* that provide information about the registration scores for a specific direction and magnitude of deviation. The registration maps from the learning set are used as input of the "Classification CNN", also based in VGG-19 architecture, that is responsible to discriminate between *genuine* and *impostor* maps. Once again, the output layer is a *soft max* loss corresponding to the probability of a set of registration maps regarding a *genuine* comparison. As previously, learning was done according to the stochastic gradient descend algorithm, with an initial learning rate of 1e^{-2}, momentum set to 0.9 and weight decay equals to $1e^{-3}$.

2.2. Classification Phase

The classification (recognition) processing chain is enclosed by the red dashed rectangle in Fig. 2. For a query+gallery pair of iris images, the detection and mapping process is carried out in an exactly equal way to the used in the "C" group of the learning phase. We use the SSD [13] detector to get information about the iris center and size $[\boldsymbol{x}_c, s]$ and obtain the *segmentation-less* polar representation of the iris. Finally, according to the overlapped grid control points, we obtain the registration maps that feed the learned classification CNN model, which returns the probability (output of a *softmax* layer) that the query and the gallery regard the same identity.

3. Results and Discussion

3.1. Experimental Setting and Preprocessing

Experiments were conducted in three datasets: CASIA-IrisV4-Lamp, CASIA-IrisV4-Thousand [1] and WVU [2], with examples of the predominant degradation factors given in Fig. 4: off-angle and partially occluded irises, glasses, extremely dilated/constricted pupils (in all sets) and local illumination variations (in WVU). 822 classes (411 eyes) from the CASIA-IrisV4-Lamp set, 2,000 classes (1,000 eyes) from the CASIA-IrisV4-Thousand and 638 classes from the WVU set were considered. For the CASIA sets, 10 images per class were used, while for the WVU set the number of images per class varied between 2 and 10. Identities were separated into two halves for learning/test (the first half of the identities, according to the alphanumeric order of the filenames, was chosen for learning), with the resulting data enabling to perform 18,495 (genuine) + 8,425,500 (impostor) pairwise comparisons for the CASIA-IrisV4-Lamp set, 45,000 (gen.) + 49,950,000 (imp.) for the CASIA-IrisV4-Thousand and 11,301 (gen.) + 7,470,767 (imp.) for the WVU set. In every experiment, the number of impostor pairwise comparisons considered was limited to 1,000,000 (randomly chosen).

Using a coarse-to-fine segmentation strategy [30], composed of a form fitting step and a geodesic active contours algorithm, images were segmented, with the iris boundaries being described by shapes of 20 degrees-of-freedom (dof) and the scleric boundary described by 3 dof. The accuracy of the segmentation step was manually verified, and different parameterizations considered, to assure that all the samples were appropriately segmented. The ambitious *open-world* setting was the unique considered, i.e., all images from half of the identities in each set were used for learning purposes, with the recognition experiments being conducted in previously unseen identities.

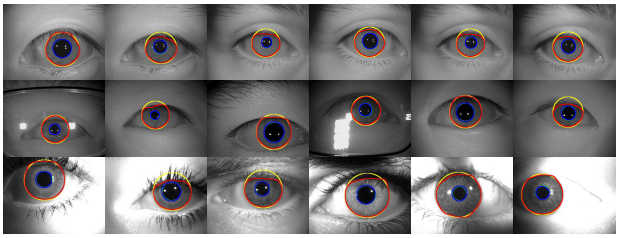


Figure 4. Datasets used in our experiments: from top to bottom rows, images of the CASIA-IrisV4-Lamp, CASIA-IrisV4-Thousand and WVU sets are shown.

As baselines, five methods were considered to represent the state-of-the-art : 1) IRINA [25], which analyses in a non-holistic way the correspondences between iris patches in the segmented and normalized irises; 2) the work due to Yang *et al*. [36] (using the O^{2}PT *iris-only* variant, with block size $w = 2, h = 14$, translation vector $[6, 3]^T$ and neighbourhood 8×8); 3) Sun and Tan's [31] method (with di-lobe and tri-lobe filters, Gaussians 5×5, $\sigma = 1.7$, inter-lobe distances {5,9} and sequential feature selection); 4) the keypoints-based method due to Belcher and Du [1] (with 64 bins = 4 (horizontal) $\times$ 4 (vertical) $\times$ 4 (orientation), SIFT descriptors extracted using VLFeat package [3]); and 5) the OSIRISv4 [22] framework, to represent the classical processing chain proposed by Daugman [5] (using the Viterbi algorithm for segmentation, Gabor filters for feature extraction and the Hamming distance for matching).

Three performance measures are reported: the decidability index (d'), the area under curve (AUC) and the equal error rate (EER), also providing the corresponding receiver operating characteristic (ROC) curves. The pairwise comparisons per dataset were divided into random samples (drew with repetition), each one with 90% of the available pairs. Then, tests were conducted iteratively in each sample, with the results considered to approximate the confidence intervals at each point, in a bootstrapping-like strategy. To facilitate the reproducibility of the experiments reported here, we release the MatConvNet [35] trained "Registration" and "Classification" VGG-19 models [4].

3.2. Parameter Tuning

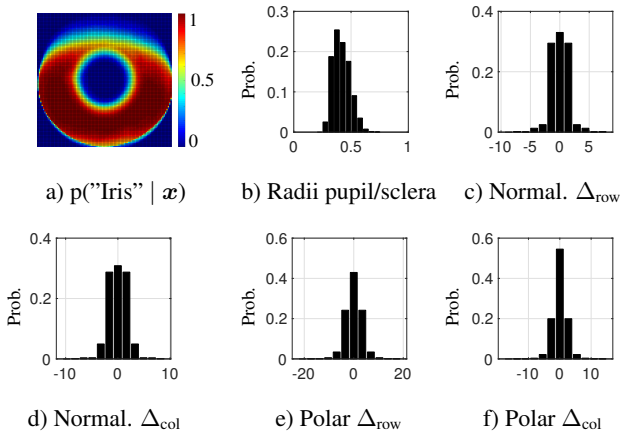
The performance of our method depends of three parameters: 1) the minimum probability τ_1 required for relative positions in the eye patches being "iris" to be considered in the analysis (p("Iris" $\mid \boldsymbol{x}) > \tau_1$, $\boldsymbol{x} = [x, y]$); 2) the size of the grid $(\tau_r \times \tau_c)$ overlapped in the polar representations of the data; and 3) the neighbourhood $\pm(\Delta_r, \Delta_c)$ considered when searching for biologically corresponding patches.

[1] CASIA iris image database, http://biometrics.idealtest.org

[2] West Virginia University iris dataset, http://www.clarkson.edu/citer/research/collections/

[3] http://www.vlfeat.org/

[4] http://tinyurl.com/y3e6l73r



a) p("Iris" | $\boldsymbol{x}$) b) Radii pupil/sclera c) Normal. Δ_{row}

d) Normal. Δ_{col} e) Polar Δ_{row} f) Polar Δ_{col}

Figure 5. a) Probability that a pixel inside the detected "Eye" region is *noise-free* and *iris*; b) Relative radii between the pupil and the sclera; c)-d) rows/columns offset between corresponding points in the normalized images (in pixels); e)-f) corresponding offsets in the *segmentation-less* polar images.

In Fig. 5 we provide the p("Iris" | $\boldsymbol{x}$) values for the learning data. For each bounding box parameterized by $[\boldsymbol{x}, r] \in \mathbb{N}^3$ (i.e., $\boldsymbol{x} \rightarrow \boldsymbol{x} + [r, r]$) we used the segmentation information to classify each position into "iris" or "non-iris". It can be seen that there is a residual probability of observing iris pixels inside a central circle (corresponding to the pupil) and at the image corners (corresponding to sclera). Also, the upper part of the iris is often occluded by eyelids/eyelashes, which reduces the "iris" probability in these regions. Setting $\tau_1 = 0$, a binary mask $\boldsymbol{B}$ can be obtained:

$$\boldsymbol{B}(\boldsymbol{x}) = \begin{cases} 1, & \text{if } p(\text{"Iris"}|\ \boldsymbol{x}) > \tau_1 \\ 0, & \text{otherwise} \end{cases}$$

Using the circular Hough transform, two concentric circumferences $(\hat{\boldsymbol{x}}_p, \hat{r}_p)$ and $(\hat{\boldsymbol{x}}_s, \hat{r}_s)$ were fitted to $\boldsymbol{B}$, and deemed to be the *generic* pupillary and scleric boundaries. These parameterizations are the key to obtain the *segmentation-less* polar representations of the iris, both in the learning and test sets. For any image in these sets, we assume that $(\hat{\boldsymbol{x}}_p, \hat{r}_p)$ and $(\hat{\boldsymbol{x}}_s, \hat{r}_s)$ correspond to the pupillary and scleric boundaries and normalize the data, according to (1).

In terms of the $(\tau_r \times \tau_c)$ values, they determine the amount of information extracted from the iris, and increase quadratically the time complexity of the encoding step. Keeping $\frac{\tau_r}{\tau_c} = \frac{h}{w} = \frac{1}{2}$, we assessed the performance of our method in a CASIA-4-Lamp validation set with $\tau_r = \{2, 4, 8, 16, 32\}$, obtaining AUC values $\{0.980, 0.986, 0.995, 0.995, 0.995\}$ that support the choice $\tau_r = 8$. Finally, the (Δ_r, Δ_c) values were tuned according to the information taken from the learning data provided in Fig. 5 b)-f). It can be seen that, comparing the relative misalignments between the *biologically corresponding* points in the normalized (c) and d)) and in the polar (e) and f)) representations of the iris data, the major differences regard the Δ_r values, i.e., polar representations tend to provide larger row misalignments than the normalized representations, justifying the $(\Delta_r = 10, \Delta_c = 5)$ choice.

3.3. Patch-Based vs. Holistic CNNs

At a first level, we compared the performance of the method proposed in this paper to the classical (holistic) ways deep learning classification strategies have been used in biometrics research (CNNs receiving pairs of samples and discriminating between the *genuine* and *impostor* comparisons). Four CNNs were considered: 1) receiving the whole ocular samples represented in the Cartesian space (denoted as CNN-Holistic; 2) receiving the cropped iris regions, in the same Cartesian space (CNN-Iris); 3) using the segmented and normalized representation of iris data (CNN-Normalized); and; 4) using the *segmentation-less* polar representations proposed in this paper (CNN-Polar). As data augmentation techniques, two label-preserving transformations were used: 1) to simulate the scale and translation samples inconsistency, patches of scale $[0.85, 0.95]$ (values drew uniformly) were randomly cropped from the learning set; and 2) as an intensity transform, we obtained the principal components of the pixels intensities in the learning data and synthesized images by adding to each pixel multiples of the largest eigenvectors, with magnitude equal to the corresponding eigenvalues [14]:

$$\mathbf{x}^{(\text{new})} = \mathbf{x}^{(\text{old})} + [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3] \left(\alpha \odot [\lambda_1, \lambda_2, \lambda_3]^T \right), \quad (9)$$

with $\odot$ denoting the element-wise multiplication, $\mathbf{v}$ and λ denoting the eigenvectors and eigenvalues of the learning data covariance matrix and $\alpha \in \mathbf{R}^3$ being randomly drew from the Gaussian $\mathcal{N}(0, 0.1)$.

The results are provided in Fig. 6 and turn clear the advantages of the non-holistic analysis when compared to its competitors. The improvements in performance with respect to the runner up strategy (invariably CNN-Normalized) were observed for all datasets, increasing the AUC values between 5% to 15%. Among the holistic CNN frameworks, using the segmented and normalized representations of the iris provided the best performance. It is interesting to note that the *segmentation-less* polar representation got the runner-up performance in the CASIA-IrisV4-Lamp and CASIA-IrisV4-Thousand datasets and even attained the best results among all holistic representations in the WVU set. In all cases, these normalized representations provide better performance than when using the cropped

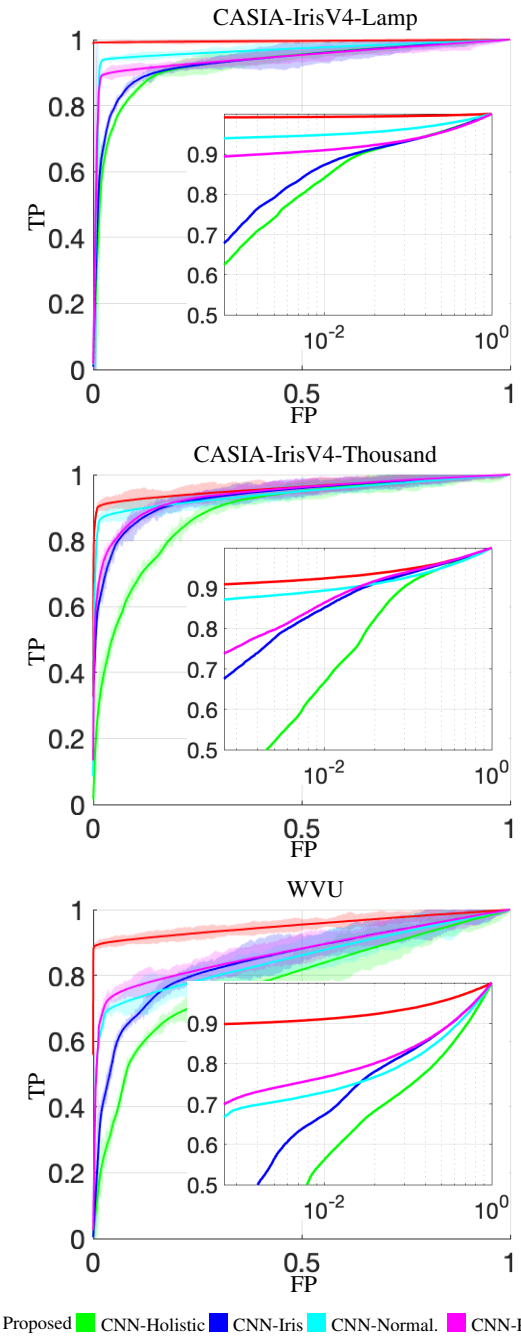


Figure 6. ROC curves of the proposed method and of four holistic CNNs typically used in biometrics: using the ocular image (CNN-Holistic); the cropped iris (CNN-Iris); the normalized iris data (CNN-Normalized); or the *segmentation-less* polar representation (CNN-Polar) as input.

eye data in the Cartesian domain and - even more evidently - than using the whole ocular region as input of the CNN.

Method	AUC	d'	EER
CASIA-IrisV4-Lamp			
Proposed Method	$0.999 \pm 5e^{-4}$	6.773 ± 0.482	0.006 ± 0.002
IRINA [25]	$0.983 \pm 8e^{-4}$	3.805 ± 0.457	0.038 ± 0.005
Sun and Tan [31]	$0.992 \pm 4e^{-4}$	4.448 ± 0.404	0.029 ± 0.005
Yang *et al.* [36]	$0.993 \pm 5e^{-4}$	4.629 ± 0.385	0.028 ± 0.004
OSIRIS [22]	$0.992 \pm 4e^{-4}$	4.017 ± 0.490	0.031 ± 0.006
Belcher and Du [1]	0.938 ± 0.007	2.933 ± 0.696	0.097 ± 0.011
CASIA-IrisV4-Thousand			
Proposed Method	$0.981 \pm 7e^{-4}$	6.611 ± 0.484	0.030 ± 0.011
IRINA [25]	$0.961 \pm 8e^{-4}$	3.600 ± 0.488	0.051 ± 0.012
Sun and Tan [31]	$0.964 \pm 1e^{-3}$	4.095 ± 0.583	0.042 ± 0.006
Yang *et al.* [36]	$0.978 \pm 7e^{-4}$	4.995 ± 0.366	0.035 ± 0.004
OSIRIS [22]	0.967 ± 0.001	3.471 ± 0.560	0.058 ± 0.007
Belcher and Du [1]	0.901 ± 0.006	2.104 ± 0.597	0.084 ± 0.011
WVU			
Proposed Method	0.969 ± 0.001	4.850 ± 0.500	0.063 ± 0.007
IRINA [25]	0.931 ± 0.002	4.066 ± 0.489	0.081 ± 0.007
Sun and Tan [31]	0.917 ± 0.002	2.552 ± 0.193	0.098 ± 0.007
Yang *et al.* [36]	0.970 ± 0.002	5.210 ± 0.185	0.055 ± 0.008
OSIRIS [22]	0.903 ± 0.005	2.801 ± 0.382	0.073 ± 0.009
Belcher and Du [1]	0.882 ± 0.011	2.008 ± 0.780	0.166 ± 0.015

Table 1. Performance summary of the proposed method with respect to three state-of-the-art methods.

3.4. State-of-the-Art Comparison

Performance comparison with respect to the baselines is shown in Fig. 7. Overall, the proposed method got either the best or the runner-up performance. Results were particularly good for the CASIA-IrisV4-Lamp, due to having learned the "Classification CNN" in (learning) pairwise samples of that set. In all cases, the approach due to Belcher and Du got the worst results, reinforcing the conclusion that keypoint-based approaches fail in iris recognition, due to the non-linear deformations that occur in the iris texture as a result of pupillary dilation. In most cases, Yang *et al.*'s performance was slightly better than ours' in the high-FAR region of the performance space. When compared to IRINA, it is evident that the method proposed in this paper represents an improvement in performance, essentially due to the use of deep-learning classification techniques to discriminate between the *genuine* and *impostors* registration maps. The open-world setting is a challenge for IRINA's matching phase, considering that the observed performance of this method was far below the reported in the original publication. Finally, Sun and Tan and OSIRIS' performance values were close to the optimal value in the highest-FAR regions of the performance space, but invariably diverged of ours and Yang *et al.*'s results in the lowest-FAR regions.

Finally, as an insight for potential ways to improve performance, in Fig. 8 we provide examples of the worst gen-

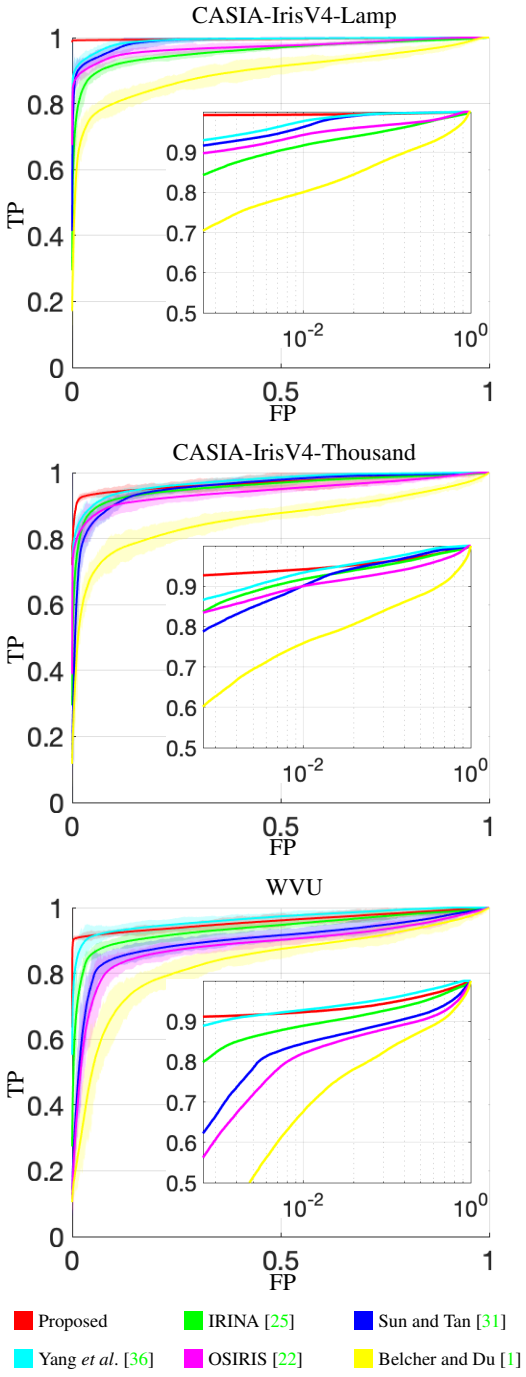


Figure 7. Comparison between the ROC curves observed for the proposed method, when compared to five techniques considered to represent the state-of-the-art.

uine (framed in green) and impostor (framed in red) comparisons, listing the factors that - under visual perception- were deemed to justify such poorest scores. Overall, the proposed method yields poor genuine scores when the bottom part of the iris is severely occluded by eyelids, whereas the worst impostor scores yielded from extremely dilated pupils that produced *segmentation-less* polar representations largely occupied by the pupillary regions.

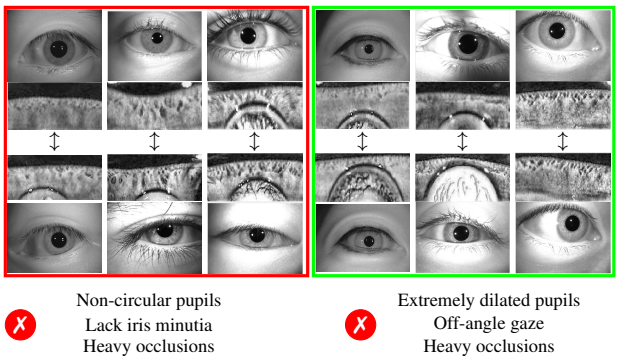


Figure 8. Example of mage pairs that produced the worst genuine (framed in green) and impostor (framed in red) matching scores in our method.

4. Conclusions and Further Work

Considering the performance of deep learning-based solutions in biometrics and the requirements of this kind of frameworks in terms of labelled data for appropriately learning *identities*, in this paper we proposed an iris recognition algorithm that uses the power of deep learning architectures exclusively to infer the misalignments between *biologically corresponding* regions in pairs of iris images represented in a *segmentation-less* polar domain. Based on the homogeneity and magnitude of the deviations, it is relatively easy to distinguish between the *genuine* and *impostor* comparisons, yielding a recognition strategy that achieves state-of-the-art performance, at a much less demanding volume of learning data than its competitors, without even requiring data to be segmented and agnostic to the levels of pupillary dilation. Our experiments turn evident that this algorithm is particularly valuable in case of low quality data, where accurate segmentation becomes a too challenging task. As main weakness, the computational complexity of the matching process should be pointed out, as the deviation maps have to be explicitly obtained for each pairwise comparison. We are now concentrated in finding alternate strategies to obtain the 2D deformation maps and reduce the computational cost of matching.

Acknowledgements

This work is funded by FCT/MEC through national funds and co-funded by FEDER - PT2020 partnership agreement under the projects UID/EEA/50008/2019 and POCI-01-0247-FEDER-033395.

References

[1] C. Belcher and Y. Du. Region-based SIFT approach to iris recognition. *Opt Lasers Engeneering*, vol. 47, no. 1, pag. 139–147, 2009. 5, 7, 8

[2] K. Bowyer, K. Hollingsworth and P. Flynn. A Survey of Iris Biometrics Research: 2008-2010. In M. J. Burge and K. W. Bowyer (eds.), *Handbook of Iris Recognition, Advances in Computer Vision and Pattern Recognition*, pag. 15–54, Springer, 2013. 1

[3] J. Chen, F. Shen, D. Chen and P. Flynn. Iris Recognition Based on Human-Interpretable Features. *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 7, pag. 1476–1485, 2016. 2

[4] J. Daugman. High Confidence Visual Recognition of Persons by a Test of Statistical Independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pag. 1148–1161, 1993. 1

[5] J. Daugman. New Methods in Iris Recognition. *IEEE Transactions on Systems, Man, and Cybernetics - part B: Cybernetics*, vol. 37, no. 5, pag. 1167–1175, 2004. 3, 5

[6] C. Ding and D. Tao. Robust Face Recognition via Multimodal Deep Face Representation. *IEEE Transactions on Multimedia*, vol. 17, no. 11, pag. 2049–2058, 2015. 1

[7] W. Dong, Z. Sun and T. Tan. Iris matching based on personalized weight map. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 9, pag. 1744–1757, 2011. 2

[8] F. Alonso-Fernandez, R. Farrugia and J. Bigun. Eigen-Patch Iris Super-Resolution for Iris Recognition Improvement. In proceedings of the 23^{rd} *European Signal Processing Conference*, doi: 10.1109/EUSIPCO.2015.7362348, 2015. 2

[9] A. Gangwar and A. Joshi. DeepIrisNet: Deep Iris Representation With Applications in Iris Recognition and Cross-Sensor Iris Recognition. In proceedings of the *International Conference on Image Processing*, pag. 2301–2305, 2016. 2

[10] S. Gao, Y. Zhang, K. Jia, J. Lu and Y. Zhang. Single Sample Face Recognition via Learning Deep Supervised Autoencoders. *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 10, pag. 2108–2118, 2015. 1

[11] S-H. Hsieh, Y-H. Li, C-H. Tien and C-C. Chang. Extending the Capture Volume of an Iris Recognition System Using Wavefront Coding and Super-Resolution. *IEEE Transactions on Cybernetics*, doi: 10.1109/TCYB.2015.2504388, 2016. 2

[12] Y. Hu, K. Sirlantzis and G. Howells. Exploiting stable and discriminative iris weight map for iris recognition under less constrained environment. In proceedings of the *IEEE International Conference on Biometrics Theory, Applications and Systems*, pag. 1–8, 2015. 2

[13] 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 proceedings of the *International Conference on Biometrics*, pag. 1–8, 2016. 2, 4, 5

[14] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In proceedings of the *Advances in Neural Information Processing Systems Conference*, pag. 1097–1105, 2012. 1, 6

[15] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C-Y. Fu and A. Berg. SSD: Single Shot Multibox Detector. arXiv:1512.02325, 2016. 3

[16] J. Long, E. Schelhamar and T. Darrell. Fully Convolutional Networks for Semantic Segmentation. In proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition*, pag. 640–651, 2015. 1

[17] N. Mahadeo, A. Paplinski and S. Ray. Optimization of Iris Codes for Improved Recognition. In proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pag. 48–55, 2014. 2

[18] M. De Marsico, M. Nappi and H. Proença Results from MICHE II: Mobile iris challenge evaluation II. *Pattern Recognition Letters*, vol. 91, no. 1, pag. 3–10, 2017. 2

[19] J. Matey, O. Naroditsky, K. Hanna, R. Kolczynski, D. Lolanoco, S. Mangru, M. Tinker, T. Zappia and W. Zhao. Iris on the move: Acquisition of images for iris recognition in less constrained environments. *Proceedings of the IEEE*, vol. 94, no. 11, pag. 1936–1947, 2006. 2

[20] D. Menotti, G. Chiachia, W. Schwartz, H. Pedrini, A. Falcão and A. Rocha. Deep Representations for Iris, Face, and Fingerprint Spoofing Detection. *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pag. 864–879, 2015. 2

[21] 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, pag. 18848–18855, 2017. 2

[22] N. Othman, B. Dorizzi and S. Garcia-Salicetti. OSIRIS: An open source iris recognition software. *Pattern recognition Letters*, vol. 82, no. 2, pag. 124-131, 2016. 5, 7, 8

[23] J. Pillai, V. Patel, R. Chellappa and N. Ratha. Secure and robust iris recognition using random projections and sparse representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 9, pag. 1877–1893, 2011. 2

[24] J. Pillai, M. Puertas and R. Chellappa. Cross-sensor iris recognition through kernel learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pag. 73–85, 2014. 2

[25] H. Proença and J. Neves. IRINA: Iris Recognition (even) in Inaccurately Segmented Data. In proceedings of the *International Conference on Computer Vision and Pattern Recognition*, pag. 6747–6756, 2017. 2, 3, 5, 7, 8

[26] H. Proença and J. 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, pag. 888–896, 2018. 1

[27] H. Qin and M. El-Yacoubi. Deep Representation-Based Feature Extraction and Recovering for Finger-Vein Verification. *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 8, pag. 1816–1829, 2017. 1

[28] A. Rodriguez, J. Panza and B. Vijaya Kumar, Segmentation-free Ocular Detection and Recognition. In proceedings of the *SPIE Conf. on Biometrics for Human Identification*, vol. 8029B, 2011. 2

[29] A. Rodriguez and B. Vijaya Kumar, Segmentation-free biometric recognition using correlation filters. *Academic Press Library in Signal Processing: Image, Video Processing and Analysis, Hardware, Audio, Acoustic and Speech Processing*, vol. 4, pag. 403–460 2013. 2

[30] S. Shah and A. Ross. Iris Segmentation Using Geodesic Active Contours. *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 4, pag. 824–836, 2009. 3, 5

[31] Z. Sun and T. Tan. Ordinal Measures for Iris Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pag. 2211–2226, 2009. 5, 7, 8

[32] C. Szegedy, A. Toshev and D. Erhan. Deep Neural Networks for Object Detection. In proceedings of the *Advances in Neural Information Processing Systems Conference*, pag. 2553–2561, 2013. 1

[33] C-W. Tan and A. Kumar. Accurate Iris Recognition at a Distance Using Stabilized Iris Encoding and Zernike Moments Phase Features. *IEEE Transactions on Image Processing*, vol. 23, no. 9, pag. 3962–3974, 2014. 2

[34] J. Thornton, M. Savvides and B.V.K. Vijaya Kumar,. A unified Bayesian approach to deformed pattern matching of iris images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pag. 596–606, 2007. 2

[35] A. Vedaldi and K. Lenc. MatConvNet: Convolutional Neural Networks for MATLAB. In proceedings of the 23^{rd} *ACM international conference on Multimedia*, pag. 689–692, 2015. 5

[36] G. Yang, H. Zeng, P. Li and L. Zhang. High-Order Information for Robust Iris Recognition Under Less Controlled Conditions. In proceedings of the *International Conference on Image Processing*, pag. 4535–4539, 2015. 2, 5, 7, 8

[37] D. Zhao, W. Luo, R. Liu and L. Yue. Negative Iris Recognition. *IEEE Transactions on Dependable and Secure Computing*, doi: 10.1109/TDSC.2015.2507133, 2016. 2

[38] Z. Zhao and A. Kumar. Towards More Accurate Iris Recognition Using Deeply Learned Spatially Corresponding Features. In proceedings of the *International Conference on Computer Vision*, pag. 3829–3838, 2017. 2

[39] Z. Zhao and A. Kumar. Improving Periocular Recognition by Explicit Attention to Critical Regions in Deep Neural Network. *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 12, pag. 2937–2952, 2018. 1

[40] M. Zhang, Q. Zhang, Z. Sun, S. Zhou and N. Ahmed. The BTAS competition on mobile iris recognition. In proceedings of the *IEEE International Conference on Biometrics Theory, Applications and Systems*, pag. 1–7, 2016. 2

[41] Q. Zhang, H. Li, Z. Sun and T. Tan. Deep Feature Fusion for Iris and Periocular Biometrics on Mobile Devices. *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 11, pag. 2897–2912, 2018. 2