

DeepGabor: A Learning-Based Framework to Augment *IrisCodes* Permanence

Hugo Proença, *Senior Member, IEEE*

Abstract—For over three decades, the Gabor-based *IrisCode* approach has been acknowledged as the *gold standard* for iris recognition, mainly due to the high entropy and binary nature of its signatures. This method is highly effective in large scale environments (e.g., national ID applications), where millions of comparisons per second are required. However, it is known that non-linear deformations in the iris texture, with fibers vanishing/appearing in response to pupil dilation/contraction, often flip the signature coefficients, being the main cause for the increase of false rejections. This paper addresses this problem, describing a customised Deep Learning (DL) framework that: 1) virtually emulates the *IrisCode* feature encoding phase; while also 2) detects the deformations in the iris texture that may lead to bit flipping, and autonomously adapts the filter configurations for such cases. The proposed DL architecture seamlessly integrates the Gabor kernels that extract the *IrisCode* and a multi-scale texture analyzer, from where the biometric signatures yield. In this sense, it can be seen as an *adaptive encoder* that is fully compatible to the *IrisCode* approach, while increasing the permanence of the signatures. The experiments were conducted in two well known datasets (CASIA-Iris-Lamp and CASIA-Iris-Thousand) and showed a notorious decrease of the mean/standard deviation values of the *genuines* distribution, at expenses of only a marginal deterioration in the *impostors* scores. The resulting decision environments consistently reduce the levels of false rejections with respect to the baseline for most operating levels (e.g., over 50% at $1e^{-3}$ FAR values). The source code of the *DeepGabor* encoder is available at: <https://github.com/hugomcp/DeepGabor>.

Index Terms—Iris recognition, Biometrics, Feature extraction.

I. INTRODUCTION

Over 30 years have passed since the appearance of the pioneering method [6] for iris recognition. During this time, the Gabor-based *IrisCode* approach has been acknowledged as the primary method for this technology, mainly due to the possibility of being used in large scale environments. When compared to competitor techniques, the main strength of this approach is its ability to effectively search in massive databases with a minimal probability of false matches, at extreme time performance. By considering binary words, pairs of signatures are matched using XOR parallel-bit logic at *lightening speed*, enabling millions of comparisons/second per processing core.

Even though various subsequent approaches have claimed to improve the recognition accuracy with respect to the baseline (e.g., [30], [25] or [18]), such methods do not work under

Author is with the IT: Instituto de Telecomunicações, Department of Computer Science, University of Beira Interior, Covilhã, Portugal, E-mail: hugomcp@di.ubi.pt

Manuscript received: May, 2022.

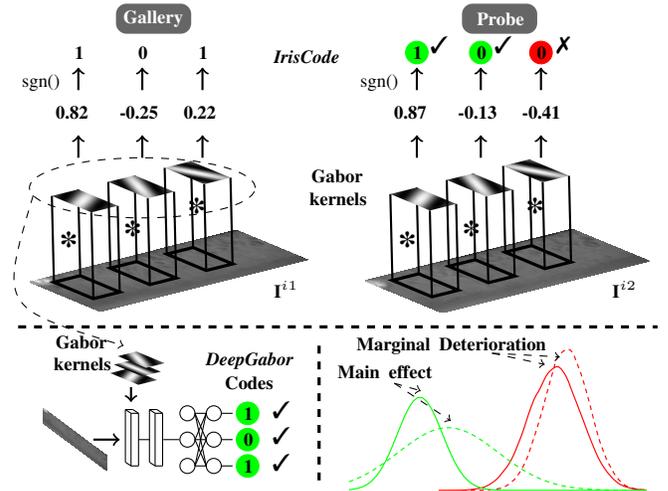


Fig. 1. Top plot: bit flipping of *IrisCode* coefficients. For a pair of images of the i^{th} subject I^{i1}/I^{i2} , non-linear deformations of the iris tissues may flip the sign of some convolution coefficients, increasing the probability of false *non-matches*. Bottom plot: this paper describes a deep learning framework that replicates the *IrisCode* bits for most cases, but also detects non-linear deformations in the iris patches and corrects (flip) the corresponding bits. As main result, the *DeepGabor* decision environments reduce the average/standard deviation values of the *genuines* distribution, at a residual deterioration of the *impostors* scores, which enables to reduce the false rejections over 50% for small FAR values ($\approx 1e^{-3}$).

the *one-shot learning* paradigm, assume multiple observations of each class to obtain appropriate decision boundaries, and - most importantly - have encoding/matching steps with time complexity that forbid their use in large environments (in particular, for *all-against-all* settings). This way, the handcrafted design of a set of Gabor filters that are further convolved to a dimensionless representation of the iris, remains as the mainstream solution.

The *IrisCode* method comprises the segmentation and normalisation phases [8] that compensate for differences in translation, scale and perspective of the acquired irises. Also, partial invariance to pupil dilation is obtained, assuming that iris deformations are linear and limited to the radial direction. However, the actual deformations in the iris texture are known to be non-linear, both radial and angular, with fibers vanishing/appearing for different levels of pupillary dilation [36]. As illustrated in the top plot of Fig. 1, such deformations often lead to changes in the sign of the convolution coefficients, *flipping* some bits of the biometric signature with respect to the gallery data. Essentially, this spread the scores of the *genuines* distribution, and increases the number of false *non-matches*,

with a corresponding degradation in users' convenience.

Aiming at increase the permanence of the signatures extracted, the main hypothesis in this paper is that the iris texture deformations that lead to bit flipping are distinguishable at the data level. Accordingly, we describe a DL-based framework that virtually emulates the *IrisCode* encoding phase and, for the large majority of the bits, reproduces the *IrisCode* coefficients. Additionally, it has the ability to detect patterns in the iris texture that lead to bit flipping, autonomously correcting the resulting coefficients for such cases.

The proposed framework can be seen as a feature encoding method that generates signatures fully compatible to the *IrisCode*, which can be matched under the same *lighting speed XOR* procedure, but are more *permanent*, in the sense that it attenuates the bit flipping problem with respect to the *IrisCode* baseline. As illustrated in the bottom-right plot of Fig. 1, when comparing the decision environments of our solution to the baseline, we notoriously reduce the mean/standard deviation of the *genuines* distribution, at expenses of a marginal deterioration of the *impostors* distribution. Overall, we consistently augment the separability between both classes and reduce the levels of false rejections for most FAR levels (e.g., over 50% at $1e^{-3}$ FAR values).

The remainder of this paper is organized as follows: Section II summarizes the most relevant research in the scope of DL-based iris recognition. Section III describes the proposed framework. In Section IV we discuss the obtained results and the conclusions are given in Section V.

II. RELATED WORK

A. Iris Recognition

Considering the maturity of the iris recognition technology, most recent strides have been concentrated in improving only particular features of the recognition process: i) extend the data acquisition volume; ii) improve the robustness to *less constrained* conditions; iii) augment the human interpretability of results; iv) develop cancellable signatures; and v) provide inter-sensor operability.

In terms of the data acquisition volume, a good example is the *iris-on-the-move* system [22], that acquires data from subjects walking through a portal. For similar purposes, Hsieh *et al.* [15] used wavefront coding and super-resolution techniques. In terms of the recognition robustness, Dong *et al.* [10] proposed an adaptive personalized matching scheme to highlight the most discriminating features. Pillai *et al.* [28] used sparse representations for classification in randomly projected iris patches, claiming to increase the robustness against acquisition artefacts. Yang *et al.* [37] relied in high-order information to perform iris matching, while Alonzo-Fernandez *et al.* [1] focused in the image enhancement phase, proposing a super-resolution method based on PCA and eigen-transformations of local iris patches. Bit consistency is also a concern, with various approaches selecting only parts of the biometric signatures for the matching step (e.g. [16], [32] and [21]).

Under complementary perspectives, the lack of interpretability hinders the application of iris recognition to forensics

domains [3]. Inter-sensor recognition provided the motivation for Pillai *et al.* [29], that learned transformations between data acquired by different sensors. Also, cancellable biometrics is a privacy-preserving solution that requires to find hardly invertible transfer functions of the biometric data between different domains: Zhao *et al.* [39] proposed the concept of negative recognition, using only complementary information (p -hidden algorithm) for matching. Finally, according to the growing popularity of CNNs, various approaches based on this paradigm appeared recently in the literature, either for specific phases of the recognition chain (e.g., segmentation [19] or spoofing detection [23]) or for the whole process [11]).

B. Deep Learning-Based Iris Recognition

It has been reported that general purpose DL-based features apply well to iris recognition: Boyd *et al.* [2] concluded that fine tuning popular models such as ResNet-50 consistently improves performance, even over models that are *learned from scratch* for the iris recognition problem. Similar conclusions were reported by Minaee *et al.* [24] and Nguyen *et al.* [26].

Yang *et al.* [38] generated multi-level spatially corresponding feature representations by means of an encoder-decoder structure, obtaining dual feature representations with complementary discriminative information. Chen *et al.* [4] addressed the large-scale recognition problem and described a loss function (tight center) that attenuates the insufficient discriminating power of the cross-entropy function. Zhao *et al.* [40] used a capsule network, with dynamic routing between layers, and three pre-trained models (VGG16, InceptionV3, and ResNet50) extracting the primary iris features. Wang and Kumar [35] introduced the concept of *residual feature* for iris recognition, describing a learning procedure with offline triplets selection and dilated convolutional kernels.

Another family of works has been concerned about the extraction of appropriate feature representations for multi-biometrics settings. Damer *et al.* [5] jointly extracted multi-biometric representations within a single DNN, creating these representations from multi-modality (face and iris), multi-instance (iris left and right), and multi-presentation (two face samples) settings. Concerned about the difficulty of performing reliable recognition in hand-held devices, Odinokikh *et al.* [27] described a lightweight CNN model that combines the advantages of handcrafted feature extractors to DL techniques. The model fuses shallow and deep feature representations to environmental features, reducing the within-subject variability.

Finally, attention mechanisms should also be highlighted: Luo *et al.* [20] used spatial and channel attention mechanisms, along with a co-attention module to obtain representative iris/periocular features. Hafner *et al.* [12] adapted the classical Daugman's pipeline, using CNNs as feature extractors and reporting performance improvements over the state-of-the-art.

III. DEEPGABOR CODES EXTRACTION

A. IrisCode: Gabor-based Feature Extraction

Gabor filters, named after Dennis Gabor¹, are linear filters used for texture analysis, that describe the frequencies content

¹https://en.wikipedia.org/wiki/Dennis_Gabor

in specific directions of image patches. Their impulse response is defined by a sinusoidal plane wave of some spatial frequency and orientation within a Gaussian envelope. Gabor filters have real and an imaginary components that represent orthogonal directions and form complex numbers:

$$\mathbf{g}(x, y, \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(\frac{x'^2 + \gamma y'^2}{-2\sigma^2}\right) \cdot \exp\left(i\left(2\pi\frac{x'}{\lambda} + \psi\right)\right), \quad (1)$$

where $x' = x \cos(\theta) + y \sin(\theta)$ and $y' = -x \sin(\theta) + y \cos(\theta)$, λ is the wavelength of the sinusoidal component, θ is the orientation of the normal to the parallel stripes of the Gabor function, ψ is the phase offset of the sinusoidal function, σ the standard deviation of the Gaussian envelope and γ is the spatial ratio that controls the ellipticity of the Gabor function.

In the iris recognition context, the convolution "*" between the dimensionless representation of the iris $\mathbf{I}(\rho, \phi)$ and a Gabor filter yields a complex-valued bit:

$$h_{Re,Im}(\rho, \phi) = \mathbf{I}(\rho, \phi) * \mathbf{g}(\rho, \phi, \lambda, \theta, \psi, \sigma, \gamma), \quad (2)$$

whose real and imaginary parts are quantized into 0/1 values, according to the $sgn(\cdot)$ function:

$$sgn(x) = \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x > 0. \end{cases} \quad (3)$$

Ignoring the coefficients with too small amplitude (considered unreliable), this process enables to encode each signature coefficient into a single bit. Note that it exclusively analyses phase information, being considered that amplitude values depend upon extraneous factors such as imaging contrast, illumination, and camera gain [7]. Even assuming that sign variations in pairs of *genuines* signatures (bit flipping) are more frequent for small magnitude coefficients, it should be noted that this phenomenon also occurs for medium/large amplitude values. Evidence about this problem is shown in Fig. 2, that provides the probability of bit flipping in the CASIA-Iris-Lamp and CASIA-Iris-Thousand for the selected 2,048 *IrisCode* bits, with respect to their magnitude. It can be confirmed that a slightly higher probability of flipping occurs for low magnitude values, but there are flipping bits across the whole interval (the higher probabilities for magnitude higher than 1 were justified to the reduced number of bits with these magnitudes).

In opposition, phase angles are invariant with respect to image contrast and focus. This way, even in cases where the phase bits are set largely on the basis of random CCD noise, this encoding strategy produces only random collisions, preventing different poorly focused irises from being confused with each other. At the end, a binary signature $\mathbf{b} = [sgn(h^1(\rho_1, \phi_1)), \dots, sgn(h^n(\rho_n, \phi_n))]$ (typically $n=2.048$) is created, with an equal number of masking bits also computed to discriminate iris regions obscured by eyelids, eyelashes, specular reflections, boundary artefacts of hard contact lenses, or poor signal-to-noise ratios.

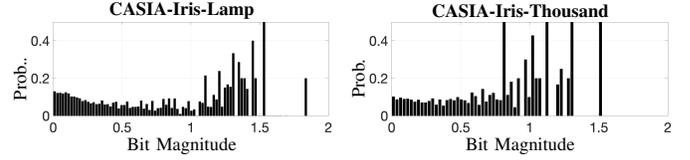


Fig. 2. Probability of bit flipping in the CASIA-Iris-Lamp (left plot) and CASIA-Iris-Thousand (right plot), with respect to the amplitude of the *IrisCode* bits. It can be seen that flipping occurs mostly - but not only - for small amplitude coefficients. The extreme probability values for large magnitude values were justified by the *small sample* problem.

B. DeepGabor: Learning Phase

Being learning-based, the framework proposed in this paper relies in a supervised set of images, annotated for identity. Assuming the availability of multiple signatures per subject, let $h^{ijk}(\rho_j, \phi_j)$ denote the k^{th} observation of the j^{th} bit (extracted from position (ρ_j, θ_j) , using the kernel configuration \mathbf{g}_j) of the signature of the i^{th} subject. We start by obtaining the modal values for all pairs subject/bit:

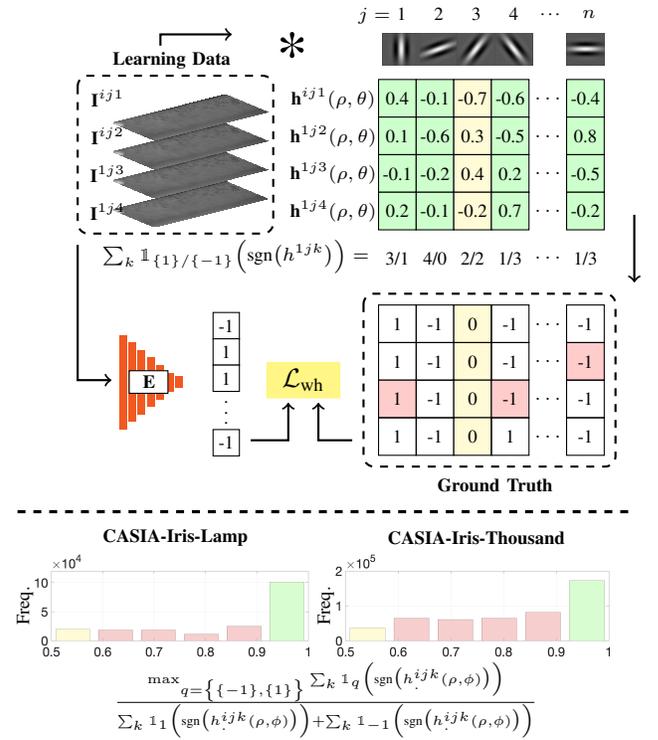


Fig. 3. Top row: cohesive idea for detecting inconsistent bits in the *IrisCode*. Using the same Gabor kernels that extract the *IrisCode*, a supervised learning set is created. Here, any code with sign different of the mode observed for a subject/bit is deemed to be *inconsistent* (in case of multi-modal observations, elements are disregarded from the learning set, in yellow). Such values are flipped and the resulting data are used as the ground-truth for the *DeepGabor* encoder **E**. Bottom row: statistics for the permanence of bits in each set.

$$h_*^{ij}(\rho_j, \theta_j) = \arg \max_{q \in \{-1, \{1\}\}} \sum_k \mathbb{1}_q \left(\text{sgn}(h^{ijk}(\rho_j, \phi_j)) \right), \quad (4)$$

with $\mathbb{1}_A : X \rightarrow \{0, 1\}$ being the indicator function of a subset A of a set X , defined as:

$$\mathbb{1}_A(x) := \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases} \quad (5)$$

The rationale for this formulation is that the modal value $h_*^{ij}(\rho, \theta)$ represents the *natural* value of that bit for a given subject, i.e., the expected value when there were no particular texture deformations in the region from where the bit was extracted. Moreover, for that subject/bit, we assume that any observations of a different value resulted from non-linear angular deformations in the iris texture, that have led to bit flipping. In these cases, the bit is considered *inconsistent* for that subject. Formally, we define an inconsistent bit/observation as:

Definition III.1. Let $h_*^{ijk}(\rho_j, \theta_j)$ represent the j^{th} bit from the k^{th} signature of the i^{th} subject. h_*^{ijk} is inconsistent iff $h_*^{ijk}(\rho_j, \theta_j) \neq h_*^{ij}(\rho_j, \theta_j)$.

Next, when creating the supervised learning set, the inconsistent bits are flipped (i.e., $\times -1$) and the corrected values are used in the supervised learning phase. This way, the DL-model receives ground-truth information that enables to associate the flipped/non-flipped values to specific patterns in the iris texture. Then, relying in the remarkable ability of deep models to distinguish between image patterns is the key for obtaining the so-called *adaptive* encoder, able to emulate the *IrisCode* or flipping some bits, depending of the textural features. Fig. 3 illustrates the whole process, where the red cells in the bottom right matrix denote the bits that were considered inconsistent and flipped. The yellow cells represent the multimodal case, where the positive/negative values have equal frequencies for a subject/bit. In such circumstances, having no other information about which is actually the *true* value, we simply disregard that bit from the learning process. The bottom row provides a statistics of the number of bits that are fully consistent (i.e., that provided constant sign for all examples/subject, in green), in opposition to bits that had a proportion of values flipped (in red), for the CASIA-Iris-Lamp and CASIA-Iris-Thousand sets. The yellow bars correspond to the number of bits that had an equal number of positive/negative values and were disregarded from the learning phase.

The resulting data are used as ground-truth of the DL-based encoder \mathbf{E} , that receives a dimensionless representation of the iris data and a set of Gabor kernels, returning the *DeepGabor* codes:

$$\hat{\mathbf{b}} = \mathbf{E}(\mathbf{I}, \mathbf{g}_1, \dots, \mathbf{g}_n). \quad (6)$$

This encoder is in practice a CNN that is tuned according to a *weighted categorical hinge* loss:

$$\mathcal{L}_{\text{wh}} = \langle \mathbf{w}, \max(0, \vec{\mathbf{1}} - \langle \mathbf{g}, \hat{\mathbf{b}} \rangle) \rangle, \quad (7)$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product between vectors, $\vec{\mathbf{1}}$ is a vector of ones, \mathbf{g} denotes the ground-truth, $\hat{\mathbf{b}}$ are the values predicted by the network and \mathbf{w} is a binary weight vector, where the 0 elements represent positions that should be ignored in computing the loss value (i.e., corresponding to

multi-modal observations).

C. DeepGabor Classification: Encoding Architecture

When designing the encoder architecture, we started by perceiving whether the COTS DL architectures, such as ResNet [13] or Inception [31], are suitable for the desired task, i.e., are able to extract in a single shot the n coefficients of the binary signature, while also flipping the values of the bits deemed inconsistent. Surprisingly, as illustrated in Fig. 4, the observed performance was very poor, as both networks invariably overfitted the feature space and completely failed to generalize. In our view, both models were trapped in local minima corresponding to spurious filter/pattern configurations in the learning sets. We hypothesize that both networks would demand a much larger amount of learning data to infer appropriate feature configurations. Due to the non-availability of such data, in practice we concluded about the non-suitability of such networks for the *DeepGabor* extraction task.

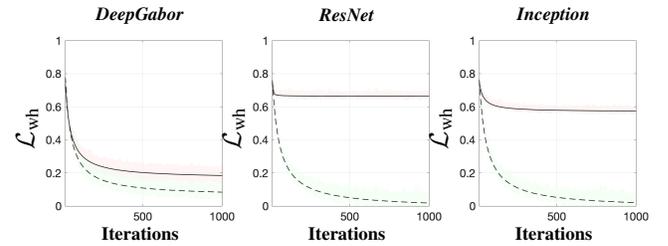


Fig. 4. Comparison between the loss values observed for the customized *DeepGabor* network model (left plot) and the values obtained for the ResNet and Inception architectures, under the *learn-from-scratch* paradigm. Results regard the CASIA-Iris-Lamp learning set, with the confidence intervals (shaded polygons) corresponding to 10 random initializations of each architecture.

Upon this conclusion, we designed a customized network architecture, tightly coupled to the requirements of this particular problem. The proposed model is depicted in Fig. 5: it receives the dimensionless representation of the iris data, along with the set of Gabor kernels (plus their position (ρ_j, θ_j) information) used to extract the *IrisCode*. The model starts by convolving the input data and the Gabor kernels, in Region I. In all convolutional layers " $s \times s, n, /i$ " denotes a convolution with n filters of size s and stride i . In parallel, a multi-scale feature encoding stage (Region II) uses kernels from $[3 \times 3]$ to $[49 \times 49]$ to obtain a detailed representation of the input data ("*Dropout* (d)" stands for a dropout layer with erasing probability equal to d). Next, all data are concatenated and feed the Deep feature encoder region (Region III), from where a global representation of the iris data is obtained (Region IV). Finally, for every bit to be extracted, the global representation is fused to a patch representation, cropped according to the (ρ_j, θ_j) position used in the *IrisCode*. Using three densely connected layers, every stream in Region V returns one *DeepGabor* bit, which are finally concatenated to obtain the biometric signature.

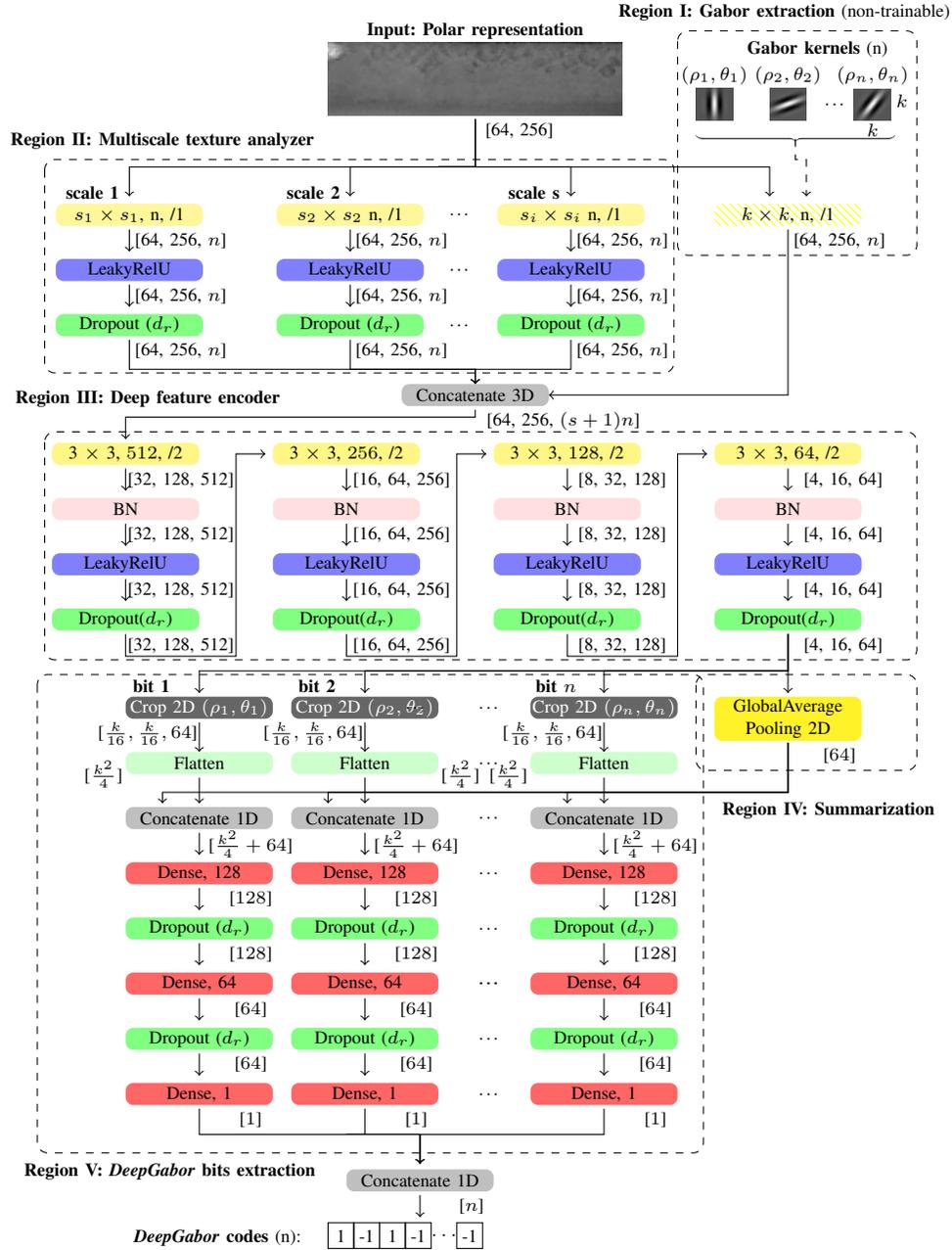


Fig. 5. Customized *DeepGabor* encoder architecture. The network analyses the dimensionless representation of the iris and returns the n -length biometric signature. Five important parts of the network are highlighted: Region I) a convolution layer which (non trainable) kernels extract the *IrisCode*; Region II) a multi-scale global feature extractor; Region III) an encoder for deeper features over global data representations; Region IV); a heavily condenser representation of the global data; and Region V) a set of local analysers, that return the *DeepGabor* codes. Notation: convolution layers appear in yellow, where $s \times s, n, /d$ represents n kernels of size s , with stride d . Dropout layers appear in green, with the dropout rate between parentheses. "BN" stands for batch normalization layers. Bi-dimensional cropping layers are denoted by "Crop 2D (x, y)", centered at position (x, y) . Finally, "Dense, n " denotes a fully connected layer, with n neurons.

IV. RESULTS AND DISCUSSION

A. Experimental Setting and Preprocessing

Our experiments were conducted in two datasets that are well known in the iris recognition context: the CASIA-Iris-Lamp and the CASIA-Iris-Thousand [33]: 1) The CASIA-Iris-Lamp was collected using a hand-held sensor, with a lamp turned on/off to augment the intra-class lighting variations. Elastic deformations of the iris texture due to pupil

dilation/contraction are present, which makes it suitable for studying problems of non-linear iris normalization and robust iris feature representation. Here, 16,212 images from 819 subjects were considered; and 2) the CASIA-Iris-Thousand, that contains 20,000 images from 1,000 subjects, collected with a IKEMB-100 camera from IrisKing². In this set, the main sources of intra-class variations are eyeglasses and

²<http://www.irisking.com>

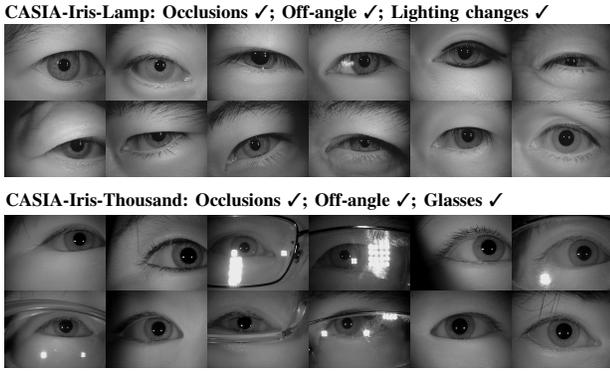


Fig. 6. Datasets used in our experiments: from top to bottom rows, images of the CASIA-Iris-Lamp and CASIA-Iris-Thousand [33] are shown. The typical data variation factors in each set are also highlighted.

specular reflections. Examples of the images in both sets are given in Fig. 6, showing the main degradation factors: off-angle and occluded irises, glasses, dilated/constricted pupils and shadows.

All images used were successfully segmented [34] and normalised into the pseudo polar domain [9], with the right halves discarded (corresponding to the upper half of the irises). The resulting data were resized to 256×64 pixels, and 10% of the subjects in each set (82 for CASIA-Iris-Lamp and 100 for CASIA-Iris-Thousand) randomly chosen for the Gabor feature selection process, to obtain the 2,048 filter configurations that extract the *IrisCode* signatures. The remaining subjects were split into two parts, with 70% of the subjects included in the learning data and the other 30% used as test subjects. In this setting, note that the “feature selection”, “learning” and “test” sets not only have disjoint images, but also disjoint subjects, which reduces the probability of overfitting.

We selected the *IrisCode* method as main baseline. Even if there is not complete information available about the actual parameters of the Gabor kernels that it uses, it is known that two multi-scale 2-D wavelet size parameters span an eight-fold range from 0.15 to 1.2 mm on the iris, and the wavelet frequency spans three octaves in inverse proportion to the size parameters [8]. Considering a set of $64 = \{8 \times 4 \times 2 \times 1 \times 1\}$ filter parameters combinations ($\lambda = \{\sqrt{2}, 2, 2\sqrt{2}, 4, 4\sqrt{2}, 8, 8\sqrt{2}, 16\} \times \theta = \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\} \times \psi = \{0, \frac{\pi}{2}\}, \sigma = \frac{\lambda}{2}, \gamma = \{1\}$), at one of 32×128 positions (ρ, ϕ) in the dimensionless representation of the iris, yields 262.144 possible features. Then, according to the sequential forward floating selection algorithm [17], we iteratively added one feature at a time to an empty set of features, until the addition of extra features didn’t improve the decidability criterion d' :

$$d' = \frac{|\mu_G - \mu_I|}{\sqrt{\sigma_G^2 + \sigma_I^2}}, \quad (8)$$

being μ_G, μ_I the mean values of the *genuines* and *impostors* distribution and σ_G, σ_I the corresponding standard deviation values. For two-choice decision tasks, such as the pairwise identity verification task that we assume in our experiments,

d' measures how well separated the two distributions are and reflects the degree to which any improvement in the false match error rate augments the false non-match error rate. Also, it is particularly suitable for working in unimodal (and roughly Gaussian) distributions, as in our case.

All our experiments were conducted according to a bootstrapping-like strategy: having n test images available, the bootstrap randomly selects (with replacement) $0.9 \times n$ images, obtaining samples composed of 90% of the whole data. Ten test samples were created, with experiments conducted independently on each trail, and obtaining the mean and the standard deviation performance among the ten trails. Obviously, classical statistical significance tests could have been used instead. However, it is known that in extremely large samples (as in our case), we’re more likely to obtain statistically significant results, even if the effect is actually small or negligible in the real world. This way, small effects can be exaggerated if they meet the significance threshold, while interesting results can also ignored when they fall short of meeting the threshold.

As comparison terms, we provide the results for: 1) IRINA [30], which is a DL-based method that analyses in a non-holistic way the correspondences between iris patches in the normalized irises; and 2) Yang et al. [37] (with FV-iris, block size $w = 2, h = 14$, translation vector $[6, 3]^T$ and neighbourhood 8×8). Regarding the performance obtained by these methods, it should be noted that the direct comparison to the *DeepGabor* and *IrisCode* methods is not fair, as their time complexity constraints the possibility of being used in large scale scenarios (in particular for duplication checks). Even if both methods start from the segmented and normalized representation of the iris data (as our solution), 1) IRINA is extremely costly in terms of the matching phase, requiring a *registration map* for every pairwise comparison to be carried out; and 2) [37], that is not as timely costly as [30], but, apart from a keypoint description phase, it also requires to train a GMM for every iris region to be matched, followed by Hamming + Euclidean distances calculus for every pairwise comparison.

To illustrate this sensitive point, note that check duplication among all the 1.2 billion people registered in India, requires about 7.2×10^{17} matching operations. Even on a 600 million people database, a duplication check on every three years, would require a system capable of more than 1.9 billion matches per second without interruption. For reference, Table I compares the average number of comparisons per second for our implementations of the four algorithms analysed (ten trials of 100,000 comparisons were considered, from where the mean \pm the standard deviation values were taken). Note that no particular attention was paid in optimizing the code, and a high level programming language running in a single CPU was used. Even though, these values provide broad evidence of the dramatically higher temporal performance of [30] (and also of [37]) with respect to *DeepGabor/IrisCode* baselines. As such, we consider a requirement that the matching step of a large scale recognition algorithm exclusively computes the Hamming distance.

The *DeepGabor* networks used 10% of the learning data for

TABLE I

AVERAGE NUMBER OF COMPARISONS PER SECOND, OBTAINED FOR OUR IMPLEMENTATIONS OF THE *DeepGabor*, *IrisCode*, IRINA [30] AND YANG *et al.* [37] METHODS.

Method	Comparisons/second
<i>IrisCode</i> , <i>DeepGabor</i>	3,842,307.02 \pm 4,027.51
IRINA [30]	2.38 \pm 0.94
Yang <i>et al.</i> [37]	404,381.15 \pm 3,012.00

validation purposes and were initialized with random weights, from zero-mean Gaussian distributions with standard deviation $1e^{-2}$ and bias 0.5. The initial learning rate was set to $1e^{-3}$, with decay $1e^{-2}$ and momentum 0.8. The learning process was stopped when no improvements in the validation loss occurred for 10 consecutive epochs. We provide the Receiver Operating Characteristic (ROC) curves, that relate the True Positive/False Positive rates, at various thresholds. Based on these plots, we provide the Area Under Curve (AUC) values, to summarize the classifier performance in a single number, and the Equal Error Rates (EER), that provide the error value when the type-I/type-II errors are approximately equal. As a complement, the decidability index values are also provided.

B. Within-Domain Recognition Performance

We considered two types of experiments. At first, the *within-domain* setting, where the *DeepGabor* framework was trained in images of the same dataset (yet of disjoint subjects) where performance was later evaluated. This can be seen as the easiest setting, where the model has access during the learning phase to the typical deformations in the iris texture under a specific environment and data acquisition protocol.

Fig. 7 compares the recognition effectiveness obtained for the classical *IrisCode* approach and the proposed *DeepGabor* method. Each row regards one dataset, with the left plots showing the ROC curves (in linear and log scale), and the right plots providing the density plots of the *genuines/impostors* distributions.

A summary of the performance in this setting is provided in the top rows of Table II, showing the AUC, decidability (d') and EER values for the four methods evaluated. When analysing the values obtained, a noteworthy observation is that the d' /AUC/EER metrics were not always perfectly correlated, which we justified by the asymmetric shape of the *genuines* distributions obtained (in particular, the right tail of the *genuines* distributions). Anyway, we resorted to provide the ensemble (d' , EER and AUC) metrics, that were considered to provide a global perspective of the methods' performance. Being obvious that [30] and [37] outperformed the *DeepGabor* and *IrisCode* approaches, it should be noted that - as above stated - both methods are not considered suitable for being deployed in large scale scenarios. Regarding the main techniques considered, a notorious improvement of *DeepGabor* over the *IrisCode* was observed in both sets. Both the *DeepGabor* and *IrisCode* approaches got better results in the CASIA-Iris-Thousand than in the CASIA-Iris-Lamp,

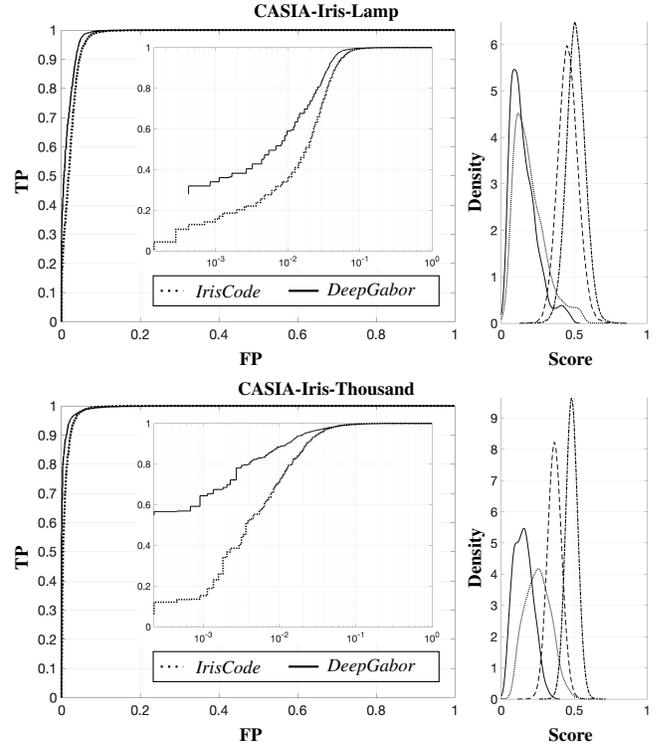


Fig. 7. Left plot: comparison between the ROC curves obtained for the baseline *IrisCode* approach and the *DeepGabor* model proposed in this paper, in the *within-domain* setting. The right plot illustrates the corresponding decision environments, where the red/green lines represent the *impostors/genuines* distributions. The dotted lines corresponds to the *IrisCode* distribution, while the solid lines represent the *DeepGabor* approach.

which did not happen for the IRINA and Yang *et al.* methods. This was justified by the features of the latter set, with more texture deformations in result of dynamic lighting conditions than the CASIA-Iris-Thousand set. Under that assumption, we concluded that IRINA and Yang *et al.* still provide more robustness to non-linear deformations in the iris texture than both the *IrisCode* and *DeepGabor* methods.

C. Cross-Domain Recognition Performance

Next, in a harder setting, we used the models learned in CASIA-Iris-Lamp to detect bit inconsistencies in the CASIA-Iris-Thousand (and vice-versa). This corresponds to a setting where not only the *optimal* filter configurations were obtained in a set of different type of the evaluation data, but also, the *DeepGabor* model should generate codes and to detect bit inconsistencies in a cross-domain setting. Fig. 8 compares the results obtained by both methods, being still notorious the advantages of *DeepGabor* with respect to the baseline. In this case, when compared to the results provided in Fig. 8, the gap in performance between the *DeepGabor* and the *IrisCode* methods was naturally smaller, but remained consistent.

D. IrisCode \leftrightarrow DeepGabor Compatibility

In our final experiment we assessed the compatibility between the *DeepGabor* and *IrisCode* signatures, which is

TABLE II

PERFORMANCE COMPARISON BETWEEN THE *DeepGabor* CODES AND THE *IrisCode* BASELINE APPROACH. ADDITIONALLY, FOR CONTEXTUALISATION, THE PERFORMANCE OBTAINED BY THE IRINA [30] AND YANG *et al.* [37] METHODS IS ALSO SHOWN.

Method	AUC	d'	EER
Within-Domain Setting			
CASIA-Iris-Lamp			
<i>DeepGabor</i>	$0.987 \pm 5e^{-4}$	2.689 ± 0.308	0.033 ± 0.003
<i>IrisCode</i>	$0.979 \pm 4e^{-4}$	2.480 ± 0.300	0.039 ± 0.002
IRINA [30]	$0.996 \pm 8e^{-4}$	7.507 ± 0.457	0.016 ± 0.003
Yang <i>et al.</i> [37]	$0.995 \pm 5e^{-4}$	4.029 ± 0.385	0.021 ± 0.003
CASIA-Iris-Thousand			
<i>DeepGabor</i>	$0.990 \pm 7e^{-4}$	3.206 ± 0.284	0.023 ± 0.004
<i>IrisCode</i>	$0.988 \pm 6e^{-4}$	2.952 ± 0.238	0.026 ± 0.003
IRINA [30]	$0.994 \pm 8e^{-4}$	6.179 ± 0.380	0.017 ± 0.004
Yang <i>et al.</i> [37]	$0.992 \pm 7e^{-4}$	3.995 ± 0.366	0.021 ± 0.004
Cross-Domain Setting			
CASIA-Iris-Thousand → CASIA-Iris-Lamp			
<i>DeepGabor</i>	$0.983 \pm 7e^{-4}$	2.401 ± 0.407	0.037 ± 0.003
<i>IrisCode</i>	$0.969 \pm 7e^{-4}$	2.350 ± 0.336	0.042 ± 0.003
IRINA [30]	$0.991 \pm 8e^{-4}$	3.800 ± 0.488	0.031 ± 0.004
Yang <i>et al.</i> [37]	$0.988 \pm 7e^{-4}$	3.659 ± 0.396	0.032 ± 0.004
CASIA-Iris-Lamp → CASIA-Iris-Thousand			
<i>DeepGabor</i>	$0.988 \pm 5e^{-4}$	2.999 ± 0.390	0.036 ± 0.003
<i>IrisCode</i>	$0.984 \pm 4e^{-4}$	2.584 ± 0.374	0.040 ± 0.002
IRINA [30]	$0.989 \pm 7e^{-4}$	3.605 ± 0.457	0.033 ± 0.005
Yang <i>et al.</i> [37]	$0.987 \pm 6e^{-4}$	3.420 ± 0.405	0.033 ± 0.004

important, considering the already deployed COTS systems. Such compatibility was tested by matching gallery and probes elements of different kinds, i.e., extracted either using the *IrisCode* approach or the *DeepGabor* networks. In practice, we aim at perceive whether it is still advantageous to use the *DeepGabor* signatures, even if they are matched against the already deployed databases of *IrisCodes*. For realism purposes, we adopted the *identification* setting in this section, and report the accumulated Rank-n values for the test data.

We started by obtaining the *mode signature* per subject, i.e., a virtual signature where each bit corresponds to the most frequent value for that bit among all signatures of the subject. Next, for each subject, the closest signature to the mode was found, by obtaining the Hamming distance between each signature and the corresponding mode, considering it as the *gallery* element. All the remaining signatures of that subject were used as probes.

Fig. 9 compares the accumulated Rank-n plots obtained in four different configurations: 1) the "Deployed Systems" group, where it is assumed that gallery elements are of *IrisCode* kind; and the "Proposed Systems" group, where the gallery elements are of *DeepGabor* kind. Inside each group, we report the performance observed when either the *IrisCode* or *DeepGabor* elements were used as the probes. Here, the first group corresponds to the most realistic scenario, i.e., use the already existing gallery data and not requiring to re-enrol the millions of subjects registered in the current systems. The

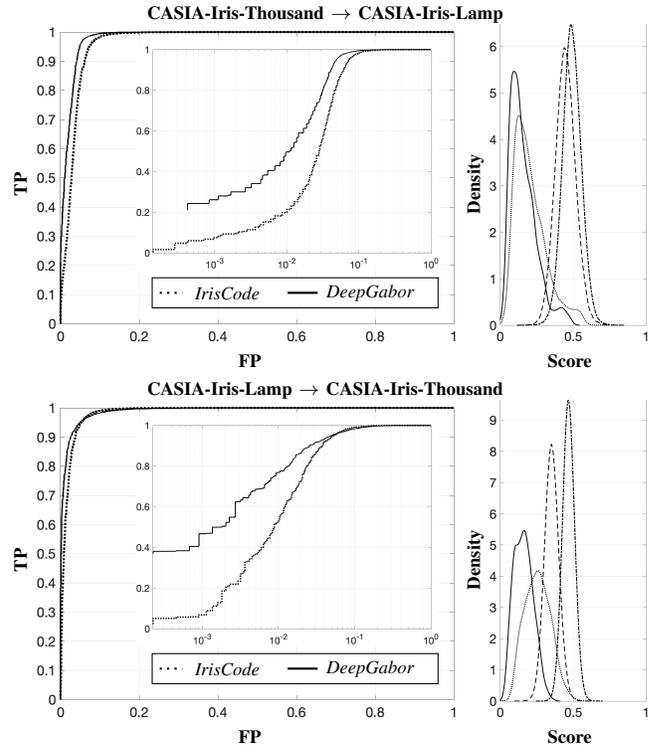


Fig. 8. Left plots: comparison between the ROC curves obtained for the baseline *IrisCode* approach and the *DeepGabor* model proposed in this paper. The right plot illustrates the corresponding decision environments, where the red/green lines represent the impostors/genuines distributions. The dotted lines corresponds to the *IrisCode* distribution, while the solid lines represent the *DeepGabor* approach.

vertical axis provides the *hit rate*, while the horizontal axis corresponds to the relative penetration ($[0-1]$), with respect to the number of subjects in the gallery. The inner plot is a zoomed-in region that provides the accumulated top-10 hit rates. Values are provided for the CASIA-Iris-Lamp (upper plot) and CASIA-Iris-Thousand (bottom plot) sets. It can be seen that - for both sets - the outperforming performance was observed for the "Proposed System" setting, when using *DeepGabor* elements both in the gallery and probe sets. Then, the runner-up values were observed for the "*IrisCode* ↔ *IrisCode*" setting. For both datasets, the "*DeepGabor* probes ↔ *IrisCode* gallery" configurations attained similar results to the "*IrisCode* ↔ *IrisCode*" settings, which corresponds to the currently deployed systems. This was considered a positive indicator for the compatibility between the proposed method and the already deployed systems. In this case, even if the recognition robustness doesn't improve evidently when changing the probes from *IrisCode* to *DeepGabor*, it is at least assured that no decreases in recognition rate happen.

E. Why Does it Work?

Finally, we hypothesise about the reasons that sustain the improvements in performance of the *DeepGabor* approach with respect to the *IrisCode*. To obtain *better* signatures than the baseline, the basic premise is that a positive balance

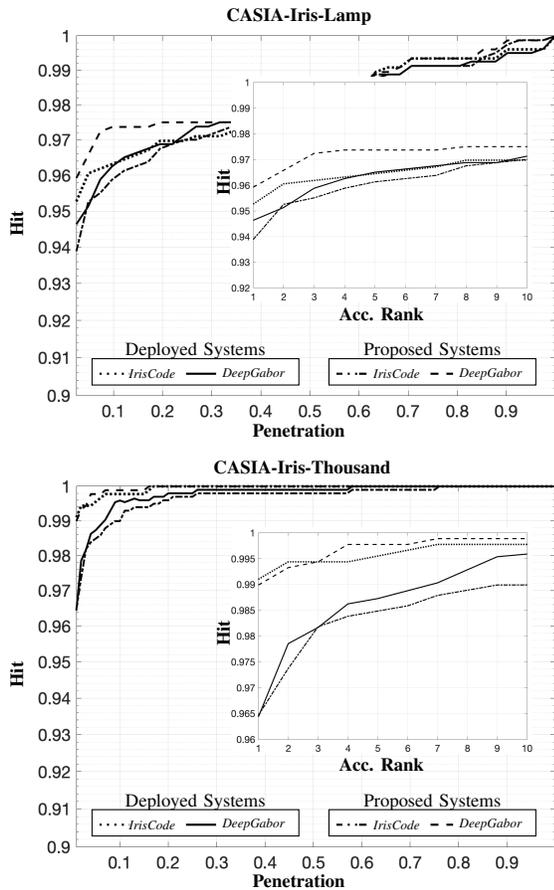


Fig. 9. Comparison between the accumulated Rank-N plots (expressed in terms of hit/penetration values), for the already deployed systems. i.e., considering the *IrisCodes* as gallery data and then using either the *DeepGabor/IrisCode* as probes (“Deployed Systems” group). Also, the corresponding values for systems with *DeepGabor* enrolled codes are shown (“Proposed Systems” group). Results are given for the CASIA-Iris-Lamp (upper plot) and CASIA-Iris-Thousand (bottom plot) sets.

between better/worse *DeepGabor/IrisCode* bits should be obtained, i.e., more *DeepGabor* bits should be *better* than their counterparts *IrisCode*, when compared to the opposite case (*IrisCode* better than *DeepGabor*). To confirm this assumption, we used the decidability index (8) and directly compared the individual discriminability of *DeepGabor/IrisCode* bits. This statistics enables to perceive the relative differences in the distributions generated for the *genuines/impostors* scores. Results are given in Fig. 10 for the CASIA-Iris-Lamp (left plots) and CASIA-Iris-Thousand sets (right plots). The upper figures provide the samples decidability values for the 2,048 bits selected per dataset, while the bottom plots provide the corresponding density estimates, according to a kernel density estimate [14]. In all plots, the horizontal axes correspond to the *IrisCode* approach, whereas its vertical counterparts regard the *DeepGabor* approach. The regions above the line $y = x$ denote bits that *improved* their decidability upon the correction by the *DeepGabor* model, while the regions below that line regard the opposite case, i.e., when the *DeepGabor* corrections even decreased the overall bit decidability. In both datasets, it is evident that the large majority of the bits improved their

decidability *IrisCode* \rightarrow *DeepGabor* (1,475 bits (72%) in the CASIA-Iris-Lamp and 1,778 bits (86.8%) in the CASIA-Iris-Thousand sets were above the $y = x$ line), which we considered to be the main justification for the improvements in the overall recognition performance.

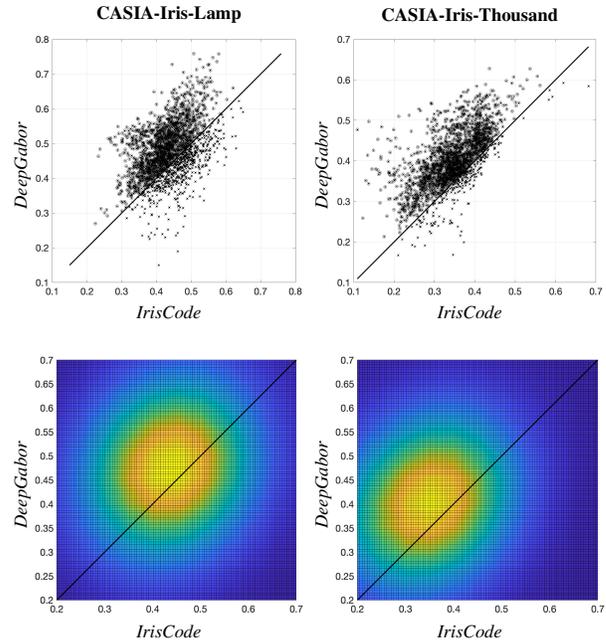


Fig. 10. Comparison between the individual decidability (8) of the 2,048 bits in the *IrisCode* (horizontal axes) and *DeepGabor* (vertical axes) approaches. Results are provided for the CASIA-Iris-Lamp (left plots) and CASIA-Iris-Thousand (right plots) sets. The plots given in the bottom row are the corresponding bivariate density estimates. The straight diagonal lines correspond to regions where the *IrisCode/DeepGabor* bits have equal decidability, i.e., are no better than the other.

V. CONCLUSIONS

Considering the popularity of the *IrisCode* approach for iris recognition and the evidence of being the *de facto* standard for this kind of technology, this paper described a DL framework to extract biometric signatures that emulate the *IrisCode* bits, while also augmenting the robustness of the resulting codes against bit flipping, as a consequence of non-linear angular deformations in the iris texture. The resulting signatures are fully compatible to their *IrisCode* counterparts, but provide decision environments that consistently reduce the levels of false rejections, with a corresponding increase in users’ convenience.

A customized CNN architecture that seamlessly integrates the Gabor filters that extract the *IrisCode* was proposed, along with a multi-scale texture analyzer and a regressor to obtain the *DeepGabor* biometric codes. The experiments were conducted in well known datasets (CASIA-Iris-Lamp and CASIA-Iris-Thousand) and pointed for consistent improvements over the baseline, both in the within-domain and cross-domain settings.

The observed improvements in performance were generally due to an obvious decrease of the mean/standard deviation values of the *genuines* distribution, at expenses of a marginal deterioration of the *impostors* scores. Overall, this augmented

the separability between both distributions and increased the decidability of the resulting decision environments. On average, false rejections decreased over 50% with respect to the baseline at most operating levels, and particularly for low FAR values ($\approx 1e^{-3}$).

ACKNOWLEDGEMENTS

This work is funded by FCT/MCTES through national funds and co-funded by EU funds under the project UIDB/50008/2020. The contribution from Caiyong Wang in providing parts of the segmented data used in the experiments is also acknowledged.

REFERENCES

- [1] F. Alonso-Fernandez, R. A. Farrugia, and J. Bigun. Eigen-patch iris super-resolution for iris recognition improvement. In *2015 23rd European Signal Processing Conference (EUSIPCO)*, pages 76–80, 2015. [2](#)
- [2] A. Boyd, A. Czajka, and K. Bowyer. Deep learning-based feature extraction in iris recognition: Use existing models, fine-tune or train from scratch? In *2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–9, 2019. [2](#)
- [3] J. Chen, F. Shen, D. Z. Chen, and P. J. Flynn. Iris recognition based on human-interpretable features. In *IEEE International Conference on Identity, Security and Behavior Analysis (ISBA 2015)*, pages 1–6, 2015. [2](#)
- [4] Y. Chen, C. Wu, and Y. Wang. Center: A novel feature extraction approach towards large-scale iris recognition. *IEEE Access*, 8:32365–32375, 2020. [2](#)
- [5] N. Damer, K. Dimitrov, A. Braun, and A. Kuijper. On learning joint multi-biometric representations by deep fusion. In *2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–8, 2019. [2](#)
- [6] J. Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1148–1161, 1993. [1](#)
- [7] J. Daugman. How iris recognition works. In *Proceedings of the International Conference on Image Processing*, volume 1, pages I–I, 2002. [3](#)
- [8] J. Daugman. How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):21–30, 2004. [1](#), [6](#)
- [9] J. Daugman. New methods in iris recognition. *IEEE Transactions on Systems, Man and Cybernetics*, 37:1167–1175, 2007. [6](#)
- [10] W. Dong, Z. Sun, and T. Tan. Iris matching based on personalized weight map. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(9):1744–1757, 2011. [2](#)
- [11] A. Gangwar and A. Joshi. Deepirisnet: Deep iris representation with applications in iris recognition and cross-sensor iris recognition. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 2301–2305, 2016. [2](#)
- [12] A. Hafner, P. Peer, Z. Emersic, and M. Vitek. Deep iris feature extraction. In *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*, pages 258–262, 2021. [2](#)
- [13] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. [4](#)
- [14] D. Hill Peter. Kernel estimation of a distribution function. *Communications in Statistics - Theory and Methods*, 14(3):605–620, 1985. [9](#)
- [15] 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*, 46(12):3342–3350, 2016. [2](#)
- [16] Y. Hu, K. Sirlantzis, and G. Howells. Exploiting stable and discriminative iris weight map for iris recognition under less constrained environment. In *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–8, 2015. [2](#)
- [17] A. Jain and D. Zongker. Feature selection: Evaluation, application, and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:153 – 158, 03 1997. [6](#)
- [18] M. B. Lee, J. K. Kang, H. S. Yoon, and K. R. Park. Enhanced iris recognition method by generative adversarial network-based image reconstruction. *IEEE Access*, 9:10120–10135, 2021. [1](#)
- [19] N. Liu, H. Li, M. Zhang, J. Liu, Z. Sun, and T. Tan. Accurate iris segmentation in non-cooperative environments using fully convolutional networks. In *2016 International Conference on Biometrics (ICB)*, pages 1–8, 2016. [2](#)
- [20] Z. Luo, J. Li, and Y. Zhu. A deep feature fusion network based on multiple attention mechanisms for joint iris-periocular biometric recognition. *IEEE Signal Processing Letters*, 28:1060–1064, 2021. [2](#)
- [21] N. Mahadeo, A. Paplinski, and S. Ray. Optimization of iris codes for improved recognition. In *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 48–55, 2014. [2](#)
- [22] J. R. Matey, O. Naroditsky, K. Hanna, R. Kolczynski, D. J. Lofacono, S. Mangru, M. Tinker, T. M. Zappia, and W. Y. Zhao. Iris on the move: Acquisition of images for iris recognition in less constrained environments. *Proceedings of the IEEE*, 94(11):1936–1947, 2006. [2](#)
- [23] 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*, 10(4):864–879, 2015. [2](#)
- [24] S. Minaee, A. Abdolrashidiy, and Y. Wang. An experimental study of deep convolutional features for iris recognition. In *2016 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pages 1–6, 2016. [2](#)
- [25] P. R. Nalla and A. Kumar. Toward more accurate iris recognition using cross-spectral matching. *IEEE Transactions on Image Processing*, 26(1):208–221, 2017. [1](#)
- [26] K. Nguyen, C. Fookes, A. Ross, and S. Sridharan. Iris recognition with Off-the-Shelf CNN Features: A deep learning perspective. *IEEE Access*, 6:18848–18855, 2017. Invited Paper. [2](#)
- [27] G. Odinokikh, M. Korobkin, I. Solomatin, I. Efimov, and A. Fartukov. Iris feature extraction and matching method for mobile biometric applications. In *2019 International Conference on Biometrics (ICB)*, pages 1–6, 2019. [2](#)
- [28] J. K. Pillai, V. M. Patel, R. Chellappa, and N. K. Ratha. Secure and robust iris recognition using random projections and sparse representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(9):1877–1893, 2011. [2](#)
- [29] J. K. Pillai, M. Puertas, and R. Chellappa. Cross-sensor iris recognition through kernel learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(1):73–85, 2014. [2](#)
- [30] H. Proença and J. C. Neves. IRINA: iris recognition (even) in inaccurately segmented data. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6747–6756, 2017. [1](#), [6](#), [7](#), [8](#)
- [31] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826, 2016. [4](#)
- [32] 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*, 23(9):3962–3974, 2014. [2](#)
- [33] C. Wang, J. Muhammad, Y. Wang, Z. He, and Z. Sun. Towards complete and accurate iris segmentation using deep multi-task attention network for non-cooperative iris recognition. *IEEE Transactions on Information Forensics and Security*, 15:2944–2959, 2020. [5](#), [6](#)
- [34] C. Wang, J. Muhammad, Y. Wang, Z. He, and Z. Sun. Towards complete and accurate iris segmentation using deep multi-task attention network for non-cooperative iris recognition. *IEEE Transactions on Information Forensics and Security*, 15:2944–2959, 2020. [6](#)
- [35] K. Wang and A. Kumar. Toward more accurate iris recognition using dilated residual features. *IEEE Transactions on Information Forensics and Security*, 14(12):3233–3245, 2019. [2](#)
- [36] Z. Wei, T. Tan, and Z. Sun. Nonlinear iris deformation correction based on gaussian model. In S.-W. Lee and S. Z. Li, editors, *Advances in Biometrics*, volume 4642, pages 780–789. Springer Berlin Heidelberg, 08 2007. [1](#)
- [37] G. Yang, H. Zeng, P. Li, and L. Zhang. High-order information for robust iris recognition under less controlled conditions. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 4535–4539, 2015. [2](#), [6](#), [7](#), [8](#)
- [38] K. Yang, Z. Xu, and J. Fei. Dualsnet: Dual spatial attention network for iris recognition. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 888–896, 2021. [2](#)
- [39] D. Zhao, W. Luo, R. Liu, and L. Yue. Negative iris recognition. *IEEE Transactions on Dependable and Secure Computing*, 15(1):112–125, 2018. [2](#)
- [40] T. Zhao, Y. Liu, G. Huo, and X. Zhu. A deep learning iris recognition method based on capsule network architecture. *IEEE Access*, 7:49691–49701, 2019. [2](#)