IRINA: Iris Recognition (even) in Inaccurately Segmented Data

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

Abstract

*The effectiveness of current iris recognition systems depends on the accurate segmentation and parameterisation of the iris boundaries, as failures at this point misalign the coefficients of the biometric signatures. This paper describes **IRINA**, an algorithm for **I**ris **R**ecognition that is robust against **INA**ccurately segmented samples, which makes it a good candidate to work in poor-quality data. The process is based in the concept of "*corresponding*" patch between pairs of images, that is used to estimate the posterior probabilities that patches regard the same biological region, even in case of segmentation errors and non-linear texture deformations. Such information enables to infer a free-form deformation field (2D registration vectors) between images, whose first and second-order statistics provide effective biometric discriminating power. Extensive experiments were carried out in four datasets (CASIA-IrisV3-Lamp, CASIA-IrisV4-Lamp, CASIA-IrisV4-Thousand and WVU) and show that IRINA not only achieves state-of-the-art performance in good quality data, but also handles effectively severe segmentation errors and large differences in pupillary dilation / constriction.*

1. Introduction

Iris recognition is a mature technology, with systems successfully deployed in domains such as border controls, computers login and national ID cards. Since the pioneer algorithm [5] proposed in 1993, a long road has been travelled in iris biometrics research [2], with two major weaknesses subsisting:

- accurate segmentation and parameterization of the iris boundaries is required to image normalisation. As most of the iris encoding / matching strategies are phase-based, failures in segmentation lead to bit shifting in the biometric signatures, with a corresponding increase of false rejections;
- false rejections also increase in case of severely dilated / constricted pupils, which cause non-linear deformations in the iris texture that are only partially compensated by the normalisation phase. Pupil movements laterally pressure the iris, with some of the fibers folding underneath others and changing texture appearance.

Note that 1) varying lighting conditions change the levels of pupillary dilation; and 2) less constrained acquisition protocols reduce data quality and make hard to accurately parameterise the iris boundaries. Hence, the *robustness* of recognition can be seen as the major concern behind the method proposed in this paper (IRINA), keeping as main goal to achieve state-of-the-art performance in good-quality data while also handling segmentation inaccuracies and non-linear texture deformations.

A cohesive perspective of IRINA is given in Fig. 1, with a processing chain divided into three phases:

1. we estimate the posterior probabilities that patches from two iris samples *correspond*, even in case of non-linear texture deformations. Starting from a learning set of manually annotated point correspondences that define convex polygons, we densely sample these regions and obtain a large number of patches considered to regard the same biological region. This information feeds a convolution neural network (CNN), that: a) explicitly discriminates between the *corresponding* and *non-corresponding* patches; and b) implicitly learns the typical iris texture deformations;

2. we infer a free-form deformation field (set of 2D vectors) that registrates pairs of samples represented in normalised coordinates. This step is formulated using a discrete Markov random field (MRF), with unary costs provided by the responses of the CNN, and pairwise costs imposing smooth solutions that penalize local gradients of the deformation field. The loopy belief propagation (LBP) algorithm [8] is used to solve the image registration problem;

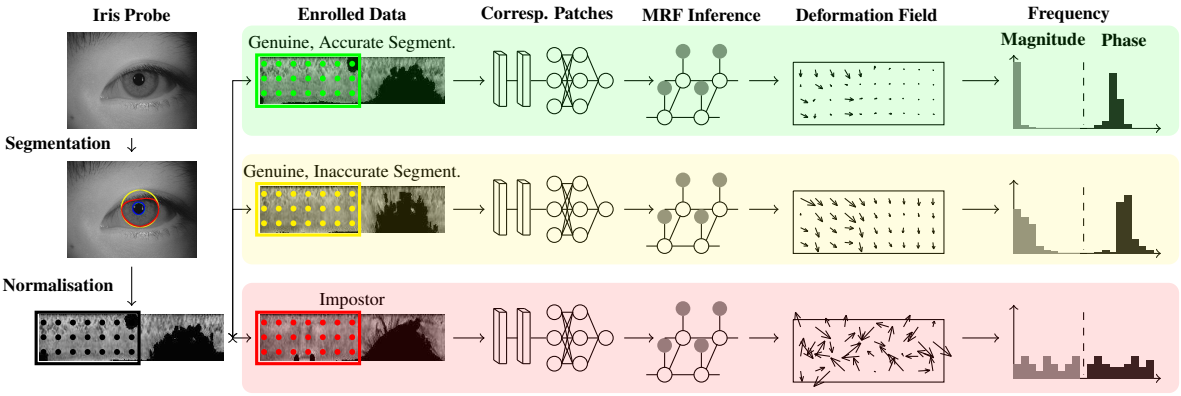

Figure 1. Overview of IRINA's processing chain. The first and second order statistics of the free-form deformation vector fields are the basis for biometric recognition. Deformation fields are exaggerated for illustration purposes.

3. for biometric recognition, the key observation is that *genuine* deformation fields (between samples of the same subject) are composed of 2D vectors with phase and magnitude gradients substantially smaller than those of *impostors*. First and second-order statistics of these vectors provide the discriminating information for biometric recognition.

Belonging to the discriminative family of pattern recognition methods, IRINA's rationale is evidently original with respect to the state-of-the-art, in which the generative paradigm rules: apart from assuming that data is accurately segmented, previous methods consider that no parts of the iris texture appear / vanish due to pupillary dilation. As an example, Thornton *et al*. [28] assume that the iris regions unaffected by pupil dilation still provide enough information for matching (providing the insight for subsequent works [14] and [24]) , while other authors provided (inevitably rough) parameterizations of iris deformations (e.g., [33], [4] and [29]). In a discriminative approach, Ross *et al*. [21] propose an information fusion framework where three distinct feature extraction and matching schemes are fused to handle the significant variability in the input ocular images. Finally, note that our idea of *corresponding* patch is different from the used in keypoint-matching iris recognition algorithms, which analyzed the geometric distribution of perfectly matching pairs of keypoints between two images (e.g., using SIFT descriptors [1]), but fail in case of varying levels of focus, lighting or non-linear iris deformations.

1.1. Iris Recognition

Given the maturity of iris biometrics technology, strides have been concentrated in improving particular features of the recognition process: i) extending the data acquisition volume; ii) improving performance in *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 [17], that acquires data from subjects walking through a portal. For similar purposes, Hsieh *et al*. [13] used wavefront coding and super-resolution techniques. In terms of the recognition robustness, Dong et al. [7] proposed an adaptive personalized matching scheme that highlights the most discriminating features. Pillai *et al*. [19] used the sparse representation for classification algorithm in randomly projected iris patches, claiming to increase the robustness against acquisition artefacts. Yang *et al*. [32] relied in high-order information to perform iris matching, while Alonzo-Fernandez *et al*. [9] 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 several approaches selecting only parts of the biometric signatures for matching (e.g. [12], [27] and [16]).

Under complementary perspectives, the lack of interpretability hinders the use of iris recognition in forensics [3]. Also, inter-sensor recognition provided the motivation for Pillai et al. [20], which learned transformations between data acquired by different sensors. Cancellable biometrics is a privacy-preserving solution that requires to find hardly invertible transfer functions of the biometric data into different domains: Zhao *et al*. [34] proposed the concept of negative recognition, using only complementary information (p-hidden algorithm) of the biometric data 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 [15] or spoofing de-

tection [18]) or for the whole process [10].

1.2. Image Registration

Image registration involves three components: i) a transformation model that maps regions of one image into regions of another; ii) a similarity criterion, that quantifies the nearness between image patches; and iii) an optimization strategy, that finds a global mapping between both images.

Transformation models can be global / local. The first family includes linear transformations such as rotation, scaling, translation and affine. Local transformations allow to warp regions of one image into another, using radial basis functions, physical continuum and large deformation models. The similarity criterion quantifies how much one image patch resembles another one in the reference data, using cross-correlation, mutual information or other distance functions. Similarity can be intensity or feature-based, with the latter family matching the most complex structures as lines and curves, based in spatial and frequency information. Finally, during optimization the set of parameters that optimally match both images are found. Exhaustive search techniques were firstly used here, but later abandoned due to their reduced computational feasibility. Modern approaches use optimization algorithms and gradient-free / gradient based techniques to derive reasonable solutions, which might be sub-optimal in case of non-convex cost functions. For additional information, Sotiras *et al*. [25] provide an overview of the state-of-the-art in image registration.

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.

2. Proposed Method

2.1. Corresponding Iris Patches

The concept of *corresponding* patches between pairs of iris images is the key to learn the typical non-linear deformations in normalized representations of the iris due to pupillary dilation / constriction and segmentation errors.

Iris recognition systems comprise a normalisation phase [6] that compensates for differences in scale, perspective and pupillary dilation, assuming that iris deformations are linear and limited to the radial direction. This does not compensate for the actual deformations, which are non-linear, radial and angular, with fibers vanishing / appearing for different levels of pupillary dilation [30]. Several authors proposed non-linear iris normalization schemes to attenuate the problem: Wyatt [31] developed a mathematical model to explain how the collagen fibers in the iris deform and Yuan and Shi [33] proposed a scheme based on that model. Also, Clark *et al*. [4] described a theoretical model for the iris dynamics, used subsequently by Tomeo-Reyes *et al*. [29].

To infer the *corresponding* patches, we use pairs of normalized samples from the same subject and manually annotate sets of control points that (by visual inspection) seem to regard the same biological region. These control points define two convex polygons $\mathbf{\Gamma}$ and $\mathbf{\Gamma}'$ and are represented by the coloured dots ($\boldsymbol{x}_i$ and $\boldsymbol{x}'_i$) in the upper part of Fig. 2. Let $\boldsymbol{x}_i = (x_i, y_i)$ and $\boldsymbol{x}'_i = (x'_i, y'_i)$, $i = \{1, \dots n\}$ be the locations of pointwise correspondences. We learn two functions f_r, f_c that establish a dense set of correspondences between positions (rows, columns) in $\mathbf{\Gamma}$ and $\mathbf{\Gamma}'$, $f_r, f_c : \mathcal{N}^2 \rightarrow \mathcal{N}$, such that $\forall\ \boldsymbol{x}'_i \in \mathbf{\Gamma}'$, $\boldsymbol{x}'_i = \big(f_c(\boldsymbol{x}), f_r(\boldsymbol{x})\big)$:

$$f_c(\boldsymbol{x}) = \boldsymbol{\lambda}_c^T [\boldsymbol{\phi}, p(\boldsymbol{x})], \tag{1}$$

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

with $\boldsymbol{\phi} = \big[\phi\big(|\boldsymbol{x} - \boldsymbol{x}_1|_2\big), \dots, \phi\big(|\boldsymbol{x} - \boldsymbol{x}_n|_2\big)\big]$, $|.|_2$ 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).

In order to obtain the $\boldsymbol{\lambda}$ coefficients, we define a $n \times n$ matrix $\boldsymbol{A}$, $A_{i,j} = \phi\big(|\boldsymbol{x}_i - \boldsymbol{x}_j|_2\big)$ and $\boldsymbol{P}$ as the $n \times 3$ polynomial basis matrix, such that $\boldsymbol{P} = [p(\boldsymbol{x}_1); \dots; p(\boldsymbol{x}_n)]$. Then, $\boldsymbol{\lambda}_c$ and $\boldsymbol{\lambda}_r$ are given by:

$$\boldsymbol{\lambda}_c = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{P} \\ \boldsymbol{P}^T & \boldsymbol{0} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{x}'_c \\ \boldsymbol{0} \end{bmatrix} \tag{3}$$

$$\boldsymbol{\lambda}_r = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{P} \\ \boldsymbol{P}^T & \boldsymbol{0} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{x}'_r \\ \boldsymbol{0} \end{bmatrix} \tag{4}$$

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

According to this procedure, we deem that positions $\boldsymbol{x} \in \mathbf{\Gamma}$ correspond biologically to $\boldsymbol{x}' = \big(f_c(\boldsymbol{x}), f_r(\boldsymbol{x})\big) \in \mathbf{\Gamma}'$. As $\mathbf{\Gamma}$ and $\mathbf{\Gamma}'$ have different size and shape, this set of correspondences implicitly encodes the non-linear deformations that affect the iris texture. Finally, we consider patches $\boldsymbol{P}$ (from $\mathbf{\Gamma}$) and $\boldsymbol{P}'$ (from $\mathbf{\Gamma}'$) of 21×21 pixels, cropped from the learning data and centered at each point correspondence.

Using 320 images (from 75 subjects) of the CASIA-IrisV3-Lamp set, 510,000 corresponding $C_{i,j}$ patches ($C_{i,j} = [\boldsymbol{P}_i, \boldsymbol{P}'_j; \boldsymbol{P}'_j, \boldsymbol{P}_i]$) were cropped. Also, using image pairs from different subjects, another 510,000 non-corresponding $\bar{C}_{i,j}$ patches were created. Both were used to train a CNN that extracts high-level texture information and distinguishes between the corresponding / non-corresponding patches. Note that the iris boundaries in this

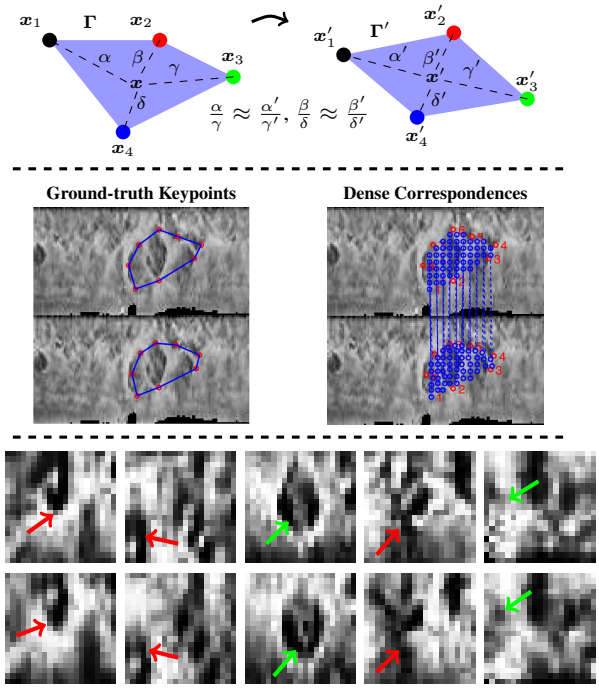

Figure 2. Concept of *corresponding* iris patches. The top part of the figure gives a schema of the way correspondences are found: based on a set of manually marked corresponding control points between two iris samples ($\boldsymbol{x}_.$ and $\boldsymbol{x}'_.$), two polygons ($\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}'$) are defined. Next, for every point inside $\boldsymbol{\Gamma}$, the corresponding position in $\boldsymbol{\Gamma}'$ is found (middle row). The bottom part of the figure shows five pairs of corresponding iris patches, where non-linear deformations (red arrows), and vanishing / emerging regions (green arrows) inside each patch are evident.

learning set (obtained as described in Sec. 3.1) were not manually confirmed, i.e., there are accurately and inaccurately samples in this set, which is important to infer the deformations in the iris texture yielding from segmentation failures.

The CNN input is 42×42 image patches and its architecture (Fig. 3) is composed of six layers (three convolutional plus three fully connected layers): the first convolutional layer uses 32 kernels (3×3), and the next ones are composed of 64 kernels of size $3 \times 3 \times 32$. The responses from these layers feed max-pooling layers (stride equals to 1, given the relatively small size of the input data). Next, there are two fully connected layers, each one with 256 cells. The output is a *soft max* loss corresponding to the probability of two iris patches to *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 responses of the CNN enable to obtain the posterior probabilities that two iris patches regard the same biological region. Such information enters a Markov random field (MRF), which energy minimization provides the solution to the image registration problem, used as information source for biometric recognition.

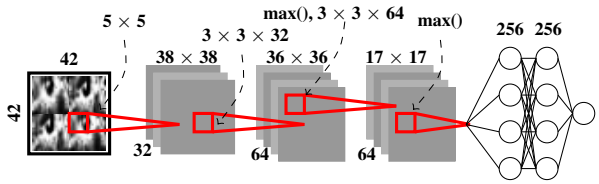

Figure 3. Structure of the convolutional neural network (CNN) used to discriminate between the *corresponding* and *non-corresponding* iris pairwise patches.

2.2. Deformation Field Inference

We consider a free-form transformation model [22] to represent a deformation field, expressed as a set of 2D vectors $\boldsymbol{d} \in \mathcal{Z}^2$ at control points $\mathring{\boldsymbol{x}} \in \mathcal{N}^2$. We superimpose a r $\times$ c regular grid at positions $\boldsymbol{G} = \{\mathring{\boldsymbol{x}}_1, \ldots, \mathring{\boldsymbol{x}}_{|G|}\}$, $|\boldsymbol{G}|$=r.c, over the left half of the normalized images representation (corresponding to the lower part of the iris that is less prune to occlusions and shadows). Also, we assume that deformations at any position $\boldsymbol{d}(\boldsymbol{x})$ can be obtained by interpolating the closest control points deformations [11]:

$$\boldsymbol{d}(\boldsymbol{x}) = \sum_{i=1}^{|G|} \nu(\boldsymbol{x})\, \boldsymbol{d}(\mathring{\boldsymbol{x}}_i), \tag{5}$$

with $\boldsymbol{d}(\mathring{\boldsymbol{x}}_i)$ representing the deformation at the ith control point and $\nu()$ being the interpolation function.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph representing a MRF, composed of a set of t_v vertices $\mathcal{V}$, linked by t_e edges $\mathcal{E}$. In our model, every control point of $\boldsymbol{G}$ is a vertex of $\mathcal{G}$, i.e., $t_v = |\boldsymbol{G}|$ and $t_e = 2.r.c - r - c$, using a typical grid configuration (4-neighborhood). The MRF is a representation of a discrete latent random variable $\boldsymbol{L} = \{L_i\}, \forall i \in \mathcal{V}$, where each element L_i takes one value l_i from a set of labels (each corresponding to a deformation vector $\boldsymbol{d}$). In practical terms, having the ith image patch centered at position $\boldsymbol{x}$, we find its corresponding patch in the second sample at positions $\boldsymbol{x}+\boldsymbol{m}$, $\boldsymbol{m}= (m_1, m_2)$, $m_i \in \{-m_{max}, \ldots m_{max}\}$. We use $m_{max} = 7$ in our experiments (Fig. 4).

Let $\boldsymbol{l} = \{l_1, \ldots, l_{t_v}\}$ be one configuration of the MRF. The energy of $\boldsymbol{l}$ is the sum of the unary $\upsilon_i(l_i)$ and pairwise $\upsilon(l_i, l_j)$ potentials:

$$E(\boldsymbol{l}) = \sum_{i \in \mathcal{V}} \upsilon_i(l_i) + \sum_{(i,j) \in \mathcal{E}} \upsilon(l_i, l_j). \tag{6}$$

According to this formulation, obtaining the deformation model between a pair of images is equivalent to infer the random variables in the MRF that minimize its energy:

$$\hat{\boldsymbol{l}} = \arg\min_{\boldsymbol{l}} E(\boldsymbol{l}), \tag{7}$$

where $\hat{\boldsymbol{l}} = \{\hat{l}_1, \ldots, \hat{l}_{t_v}\}$ ($\hat{l}_i \equiv \boldsymbol{d}_i$) are the labels inferred. In all cases, MRFs were optimized according to the Loopy Belief Propagation [8] algorithm. Even though it is not guaranteed to converge to global minimums on loopy non-sub modular graphs (such as ours), we concluded that the algorithm provides acceptable solutions most of the times.

2.2.1 Unary Costs

Let $\eta(i,j) : \mathbb{N}^2 \rightarrow [0,1]$ be the CNN response for one pair of patches, expressing the likelihood $p\big(\eta(i,j) \mid C_{i,j}\big)$ that the ith patch of one sample corresponds to the jth patch of its counterpart. According to the Bayes rule, and assuming equal priors, the posterior probability functions are given by:

$$p\big(C_{i,j} \mid \eta(i,j)\big) = \frac{p\big(\eta(i,j) \mid C_{i,j}\big)}{\sum_{k=1}^{|M|} p\big(\eta(i,k) \mid C_{i,k}\big)}, \tag{8}$$

with $|M|$ expressing the number of positions in the second image where we search for the position corresponding to the ith patch. This way, the unary costs of the labels in each vertex are defined as:

$$\upsilon_i(l_i) = \alpha \left(1 - p\big(C_{i,j} \mid \eta(i,j)\big)\right), \tag{9}$$

with $\alpha \in [0,1]$ determining the trade-off between the strength of the unary to the pairwise costs in MRF optimization.

2.2.2 Pairwise Costs

In our model, the pairwise costs serve to control the derivatives in the deformation field, i.e., penalise adjacent positions with dramatically different deformation vectors that are not biologically plausible.

Let l represent a deformation vector $\boldsymbol{d} \in \mathcal{Z}^2$ for one control point. For computational purposes, it is important to discretise the solution space, not only limiting the maximum displacement m_{max} allowed for $\boldsymbol{d}$, but also defining an appropriate sampling strategy (dense sampling produces $(2.m_{max}+1)^2$ labels). Based in [11], we use a circular sparse grid with $\frac{\sqrt{2}}{4}.\pi.\ r$ nodes, $r = \{1, \ldots, m_{max}\}$ at positions $x = i_r.\cos(\theta)$, $y = i_r.\sin(\theta)$, $\theta \in [0, 2\pi]$, being i the sampling rate at the r-radius circumference. This sparse sampling strategy reduces over 50% the number of labels without significant decreases in the method performance (leftmost part of Fig. 4).

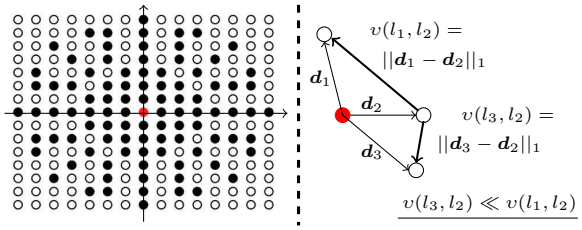

Figure 4. At left: comparison between the number of labels (maximum displacement $m_{max} = \pm 7$) when using a sparse sampling strategy, with respect to the dense sampling variant (solid black points denote the displacements $\boldsymbol{d}_i$ using the sparse sampling strategy, while the white points would have been also considered by the dense sampling strategy). At right: schema of the pairwise cost $\upsilon(l_i, l_j)$ for observing two displacement vectors $(\boldsymbol{d}_1, \boldsymbol{d}_2)$ and $(\boldsymbol{d}_3, \boldsymbol{d}_2)$ in adjacent positions of the deformation field: $\boldsymbol{d}_1$ being farther than $\boldsymbol{d}_3$ from $\boldsymbol{d}_2$ implies that $\upsilon(l_3,l_2) \ll \upsilon(l_1,l_2)$.

Finally, the pairwise cost for labelling two adjacent nodes is defined by:

$$\upsilon(l_i, l_j) = (1-\alpha)\,|\boldsymbol{d}_i - \boldsymbol{d}_j|_1, \tag{10}$$

being $|.|_1$ the ℓ_1 norm.

2.3. Classification

The biometric recognition task is regarded as a binary classification problem. We use a machine-learned classifier to discriminate between the set of features extracted from positive (*genuine*) and negative (*impostor*) pairwise deformation fields. Let $\hat{\boldsymbol{l}} = \{\hat{l}_1, \ldots, \hat{l}_{t_v}\}$ represent the set of labels returned by the MRF. Each label l_i corresponds bijectively to a free form deformation vector $\boldsymbol{d}_i \in \mathcal{Z}^2$ at a position $\boldsymbol{x}$ of the normalised coordinates space. We extract the histogram of magnitudes and phase angles of $\boldsymbol{d}_i$ and their second-order statistics (local energy and homogeneity) from the magnitude and phase maps (with 6×12 vectors, taken in 3×3 and 5×5 regions, using stride 3 and 5), yielding 34 features that feed the binary discriminant (SVM in our case). A disjoint set from the CASIA-IrisV3-Lamp set (with 3,000 genuine / 3,000 impostor pairwise comparisons) was used as learning data at this point.

3. Results and Discussion

3.1. Datasets and Experimental Setting

IRINA was empirically validated in four iris datasets: CASIA-IrisV3-Lamp, CASIA-IrisV4-Lamp, CASIA-

IrisV4-Thousand [1] and WVU. [2]. Examples are given in Fig. 5, showing the degradation factors of each set: off-angle and occluded irises, glasses, dilated / constricted pupils (all sets) and shadows (WVU). 500 classes (eyes) per data set were used: for all the CASIA-Iris sets, 10 images per class were considered, while for the WVU the number of images per class varied between 2 and 10. All images were successfully segmented according to a coarse-to-fine strategy [23], composed by a form fitting step and a geodesic active contours algorithm. This way, we accurately parameterize the iris boundaries, having the pupillary contour described by shapes of 20 degrees-of-freedom (dof) and the scleric boundary described by 3 dof. At this point, images were normalised into the pseudo polar domain [6] and their right halves were discarded (corresponding to the upper half of the irises in the original representation).

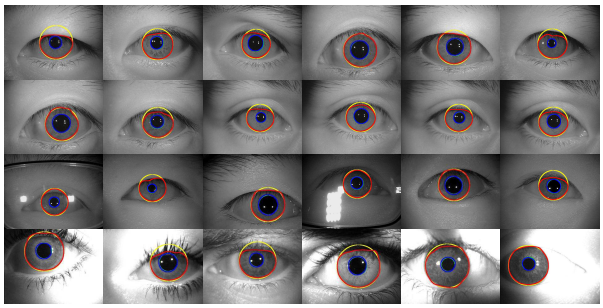

Figure 5. Datasets used in IRINA's empirical validation. From top to bottom rows, images of the CASIA-IrisV3-Lamp, CASIA-IrisV4-Lamp, CASIA-IrisV4-Thousand and WVU sets are shown.

As baselines, the methods due to Yang *et al.* [32] (using the O^{2}PT *iris-only* variant, with block size $w = 2, h = 14$, translation vector $[6, 3]^T$ and neighbourhood 8×8) and Sun and Tan [26] (with dilobe and trilobe filters, Gaussians 5×5, $\sigma = 1.7$, inter-lobe distances $\{5,9\}$ and sequential feature selection) were firstly considered, as both concern about the robustness of recognition to pupillary dilation and to non-linear iris deformations. Also, the 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]) was chosen due to the fact of being keypoints-based, even though its results cannot be considered state-of-the-art anymore. Three performance measures are reported: the decidability index (d'), the area under curve (AUC) and the receiver operating characteristic curve (ROC). In all experiments, the pairwise comparisons per dataset were divided into random samples (drew with repetition), each one with 90% of the available pairs. Then, independent performance tests were conducted in each subset, with the obtained results approximating the confidence intervals at each point, according to a bootstrapping-like strategy.

[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/

3.2. Learning and Parameter Tuning

It is important to note that the learning data used in the CNN was exclusively composed of CASIA-IrisV3-Lamp images. Using randomly sampled learning / validation and test sets (with 60% / 20% / 20% of the available pairwise comparisons), performance was tuned and all parameters strictly kept for the remaining datasets, meaning that the CASIA-Iris-V4-Lamp, CASIA-IrisV4-Thousand and WVU were used exclusively as test sets. The left plot in Fig. 6 shows the decision environment resulting from the responses $\eta(i, j)$ of the CNN, to distinguish between the *corresponding* $C_{i,j}$ and *non-corresponding* $\bar{C}_{i,j}$ iris patches. The likelihood functions $p\big(\eta(i,j) \mid C_{i,j}\big)$ and $p\big(\eta(i,j) \mid \bar{C}_{i,j}\big)$ in the CASIA-IrisV3-Lamp test set are shown.

In terms of IRINA's parameterisation, the value set to α (9) is the most sensitive, as it expresses the relative weight in the MRF between the unary and the pairwise costs. Here, $\alpha = 1$ corresponds to deformation vectors that are independent of their neighbours (no MRF would be required). In opposition, small α values reduce the local variations in the deformation field, with values below 0.9 imposing constant deformation fields with poor biometric discriminability. The AUC values obtained with respect to the value of α are shown in the right plot of Fig. 6. Note the significantly best performance (and smallest variance) for the CASIA-IrisV3-Lamp among all sets, due to the learning data that fed the CNN (same set, yet with disjoint instances). Based on these results, $\alpha = 96.35$ was used in all our subsequent experiments.

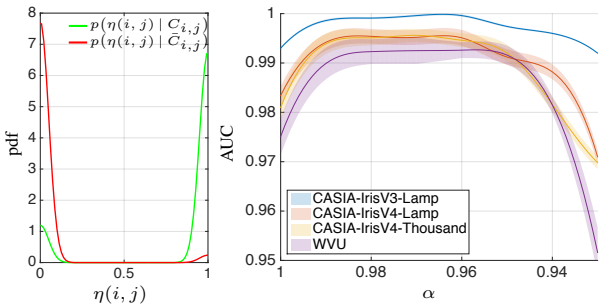

Figure 6. Left plot: decision environment of the responses given by the CNN to distinguish between *corresponding* $C_{i,j}$ and *non-corresponding* $\bar{C}_{i,j}$ iris patches (CASIA-IrisV3-Lamp set). Right plot: variations in recognition performance with respect to the α parameter.

3.3. Accurately Segmented Data

Performance was evaluated in two different settings: at first we used the accurate parameterisations of the iris boundaries, to perceive IRINA's performance in relatively good quality data. Results are given in Fig. 7, comparing the ROC curves (in linear and log scales) for the four methods and four data sets considered. It can be seen that IRINA outperformed its competitors in all cases and regions of the performance space, with exception to a narrow band around FAR $\approx 10^{-3}$ in the CASIA-V4-Thousand. In the remaining cases, IRINA was considerably better than the other methods, at some operating points with reductions in FAR levels over 40% with respect to the second best approach (usually Yang *et al.*). At the other extreme, the method due to Belcher and Du got consistently the worst results in our experiments, due to the difficulties in finding exact keypoint correspondences between images with different levels of focus or pupillary dilation. Overall, IRINA's best performance among all methods is particularly evident in the CASIA-IrisV3-Lamp, where the decreases in the error rates (over the second-best strategy) almost reached one order of magnitude.

The most relevant performance indicators are summarised in Table 1. It should be noted that results reported here should not be directly compared to the last generation of iris recognition evaluation initiatives (International Biometric Group evaluation FRR 2-5%@ FAR $\approx$ 1e^{-6} and Iris Challenge Evaluation FRR 1-3% @ FAR $\approx$ 1e^{-3}), as the average quality of data here is substantially lower than in those contests. Even though, in order to provide easy baselines, IRINA obtained FRR levels at FAR $\approx [1e^{-1}, 1e^{-2}, 1e^{-3}, 1e^{-4}]$ of [0.001, 0.006, 0.021, 0.121] (CASIA-IrisV3-Lamp), [0.012, 0.039, 0.076, 0.084] (CASIA-IrisV4-Lamp), [0.011, 0.054, 0.140, 0.156] (CASIA-Iris-V4-Thousand) and [0.023, 0.080, 0.116, 0.121] (WVU).

Method	AUC	d'	EER
CASIA-IrisV3-Lamp			
IRINA	0.999 $\pm 1e^{-4}$	12.623 $\pm$ 0.716	0.006 $\pm$ 0.001
Yang *et al.*	0.995 $\pm 4e^{-4}$	4.085 $\pm$ 0.590	0.021 $\pm$ 0.004
Sun and Tan	0.989 $\pm 5e^{-4}$	3.239 $\pm$ 0.501	0.044 $\pm$ 0.004
Belcher and Du	0.930 $\pm$ 0.005	2.701 $\pm$ 0.799	0.083 $\pm$ 0.009
CASIA-IrisV4-Lamp			
IRINA	0.995 $\pm$ 0.002	6.623 $\pm$ 0.454	0.026 $\pm$ 0.005
Yang *et al.*	0.993 $\pm 5e^{-4}$	3.629 $\pm$ 0.385	0.028 $\pm$ 0.004
Sun and Tan	0.992 $\pm 4e^{-4}$	3.448 $\pm$ 0.404	0.029 $\pm$ 0.005
Belcher and Du	0.948 $\pm$ 0.007	2.933 $\pm$ 0.696	0.077 $\pm$ 0.011
CASIA-IrisV4-Thousand			
IRINA	0.996 $\pm$ 0.001	6.179 $\pm$ 0.380	0.030 $\pm$ 0.005
Yang *et al.*	0.988 $\pm 6e^{-4}$	2.995 $\pm$ 0.366	0.045 $\pm$ 0.004
Sun and Tan	0.984 $\pm 6e^{-4}$	3.097 $\pm$ 0.583	0.052 $\pm$ 0.006
Belcher and Du	0.901 $\pm$ 0.009	2.104 $\pm$ 0.597	0.097 $\pm$ 0.012
WVU			
IRINA	0.991 $\pm$ 0.002	5.179 $\pm$ 0.361	0.042 $\pm$ 0.008
Yang *et al.*	0.980 $\pm$ 0.001	2.552 $\pm$ 0.185	0.065 $\pm$ 0.008
Sun and Tan	0.967 $\pm$ 0.001	2.210 $\pm$ 0.193	0.098 $\pm$ 0.007
Belcher and Du	0.882 $\pm$ 0.011	2.008 $\pm$ 0.780	0.116 $\pm$ 0.015

Table 1. Comparison between the performance obtained by IRINA with respect to three other strategies.

3.4. Inaccurately Segmented Data

At a second stage, we added two type of errors (translation and scale) to the iris boundaries parameterisations, to perceive the decreases in performance when the iris is inaccurately segmented. Segmentation errors of magnitude up to 21% were randomly generated, with "*magnitude*" expressing the difference between the maximum Euclidean distance between boundary points in the original and in the inaccurate segmentation parameterisation (e.g., for a circular boundary with diameter of 100 pixels, a scale error of magnitude 10% will either change the diameter to 90 or 110 pixels, whereas a translation error will move the boundary 10 pixels in a random direction).

According to our observations, the inaccurate segmentation setting is exactly when the advantages of IRINA with respect to the state-of-the-art are the most evident. The key insight IRINA' s robustness to segmentation failures is illustrated in Fig. 8, showing the deformation fields for genuine image pairwise comparisons, with accurate (green boundaries) and inaccurate (red boundaries) segmentations in the a)-c) rows, and one impostor comparison (bottom row) from the CASIA-IrisV4-Thousand set. Note that the impostor deformation field is almost chaotic, with much larger local derivatives than any genuine deformation field, where local correlation is evident.

The average decreases in performance with respect to segmentation inaccuracies up to 21% are given in Fig. 9 (mean AUC values, with 95% confidence intervals). It can be seen that IRINA almost kept its performance up to segmentation inaccuracies of 12%, and then slightly decreased its results, which could even be attenuated if larger magnitudes in the deformation field m_{max} were tolerated (yet, this would have increased the number of labels in the MRF and the computational cost). In opposition, both Yang *et al.* and Sun and Tan showed substantial decreases in performance even for relatively small segmentation errors, and almost loose any efficiency for errors larger than 15%. Finally, as it is not phase-based, the method due to Belcher and Du proved to be relatively robust against segmentation inaccuracies, but at much lower performance levels than IRINA.

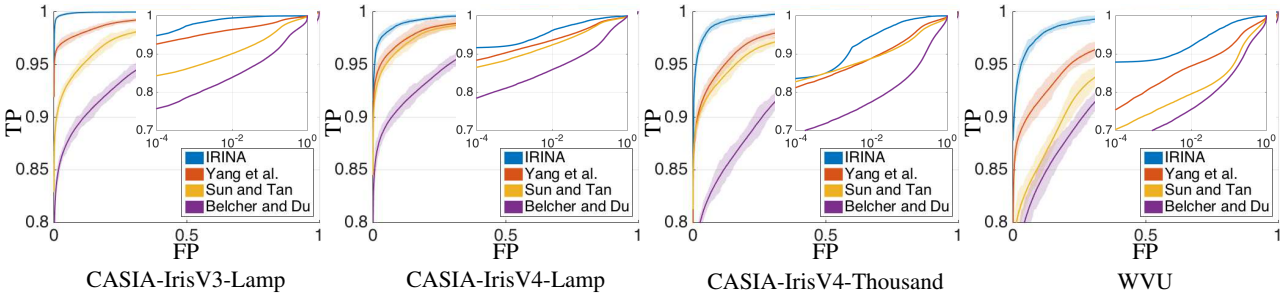

Figure 7. Comparison between the ROC curves obtained for the three methods and four datasets considered. At each operating point, the confidence interval is denoted by the shade region.

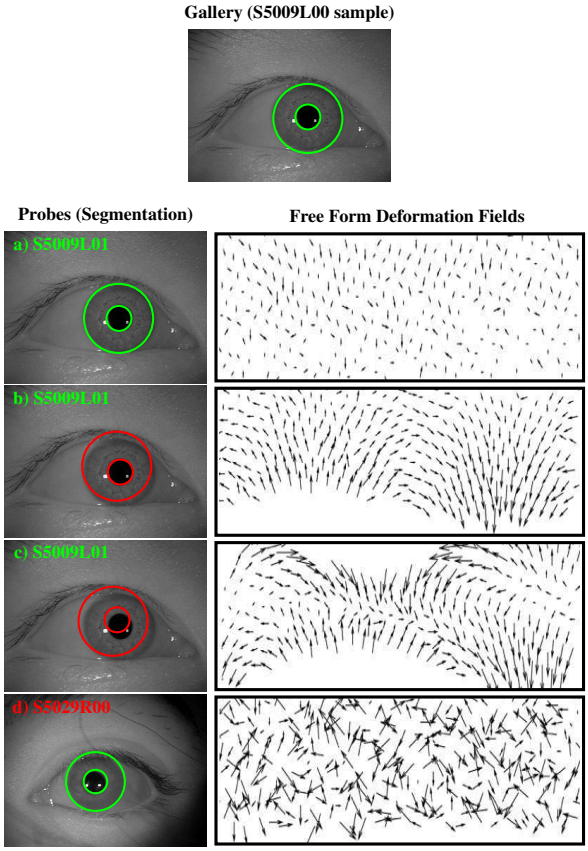

Figure 8. Examples of deformation fields with respect to failures in the segmentation of the iris. a) genuine comparison using an accurately segmented image; b) and c) genuine pairwise comparisons in inaccurately segmented data; d) impostor pairwise comparison. For illustration purposes, circular boundaries (3 dof) are used, as they provide the most evident patterns in the deformation fields.

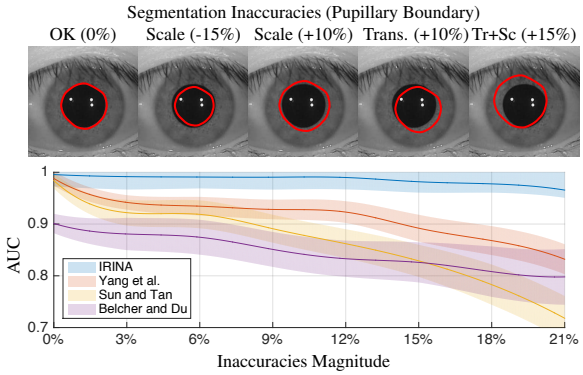

Figure 9. Top row: examples of segmentation inaccuracies and their corresponding magnitudes. Bottom row: AUC values with respect to the magnitude of segmentation inaccuracies (CASIA-IrisV4-Thousand set).

4. Conclusions and Further Work

Iris recognition has limited robustness against failures in segmentation and dilated / constricted pupils. In this paper we proposed an algorithm (IRINA) to cope with such covariates. The idea is to learn the "*corresponding*" patches between pairs of iris samples non-linearly deformed due to segmentation failures and to pupillary dilation. A convolutional neural network is used for such purposes, which provides the information for an image registration step that matches patches of the query iris sample into the enrolled data. A Markov random field infers a free form deformation field (set of 2D vectors), which first and second order statistics provide the discriminating information for biometric recognition. Our experiments show that IRINA not only achieves state-of-the-art performance in good quality data, but also effectively handles severe segmentation errors and large differences in pupillary dilation / constriction.

As current work, we are concentrated in finding alternate strategies to obtain the 2D deformation fields and reduce the computational cost of matching.

Acknowledgements

This work was supported by UID/EEA/50008/2013 research program.

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. 2, 6

[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] A. Clark, S. Kulp, I. Herron and A. Ross. A theoretical model for describing iris dynamics. In M. J. Burge and K. W. Bowyer (eds), *Handbook of Iris Recognition, Advances in Computer Vision and Pattern Recognition*, pag. 129–150. Springer, 2013. 2, 3

[5] 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

[6] J. Daugman. How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pag. 21–30, 2004. 3, 6

[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] P. Felzenszwalb and D. Huttenlocher. Efficient Belief Propagation for Early Vision. *International Journal of Computer Vision*, vol. 70, no. 1, pag. 41–54, 2006. 1, 5

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

[10] 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. 2301–2015, 2016. 3

[11] B. Glocker, N. Komodakis, N. Navab, G. Tziritas and N. Paragios. Dense Registration with Deformation Priors. *Information Processing in Medical Imaging*, vol. 21, pag. 540–551, 2009. 4, 5

[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] 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

[14] B. Vijaya Kumar, J. Thornton, M. Savvides, V. Boddeti and J. Smereka. Application of Correlation Filters for Iris Recognition. *Handbook of Iris Recognition* , pag. 337–354, 2013. 2

[15] 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

[16] 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

[17] 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

[18] 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. 3

[19] 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

[20] 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

[21] A. Ross, R. Jillela, J. Smereka, V. N. Boddeti, B. V. K. VijayaKumar, R. Barnard, X. Hu, P. Pauca and R. Plemmons. Matching Highly Non-ideal Ocular Images: An Information Fusion Approach. In proceedings of the *5th IAPR International Conference on Biometrics*, doi: 10.1109/ICB.2012.6199791, 2012. 2

[22] T. Sederberg and S. Parry. A Free-form deformation of solid geometric models. In proceedings of the *13th annual conference on Computer graphics and iterative techniques*, pag. 151–160, 1986. 4

[23] 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. 6

[24] J. Smereka, V. Boddeti and B. V. K. Vijaya Kumar: Probabilistic Deformation Models for Challenging Periocular Image Verification. *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 9, pag. 1875–1890, 2015. 2

[25] A. Sotiras, C. Davatzikos and N. Paragios. Deformable Medical Image Registration: A Survey. *IEEE Transactions on Medical Imaging*, vol. 32, issue 7, pag. 1153–1190, 2013. 3

[26] Z. Sun and T. Tan. Ordinal Measures for Iris Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pag. 221–2226, 2009. 6

[27] 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

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

[29] I. Tomeo-Reyes, A. Ross, D. Antwan and C. Vinod. A biomechanical approach to iris normalization. In proceedings of the *International Conference on Biometrics*, pag. 1–9, 2015. 2, 3

[30] Z. Wei, T. Tan and Z. Sun. Nonlinear iris deformation correction based on Gaussian model. In proceedings of the *Advances in Biometric Person Authentication, Lecture Notes in Computer Science*, vol. 4642, pag. 780–789, 2007. 3

[31] H. Wyatt. A minimum wear-and-tear meshwork for the iris. *Vision Research*, vol. 40, pag. 2167–2176, 2000. 3

[32] 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, 6

[33] X. Yuan and P. Shi. A non-linear normalization model for iris recognition. In *Advances in Biometric Person Authentication, Lecture Notes in Computer Science*, vol. 3781, pag. 135–141, 2005. 2, 3

[34] 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