Simultaneous Iris and Periocular Region Detection Using Coarse Annotations

Diego R. Lucio*, Rayson Laroca*, Luiz A. Zanlorensi*, Gladston Moreira†, David Menotti*
*Laboratory of Vision, Robotics and Imaging, Federal University of Paraná, Curitiba, PR, Brazil
†Laboratory of Intelligent Systems Computation, Federal University of Ouro Preto, Ouro Preto, MG, Brazil
*{drlucio, rblsantos, lazjunior, menotti}@inf.ufpr.br †gladston@iceb.ufop.br

Abstract—**In this work, we propose to detect the iris and periocular regions simultaneously using coarse annotations and two well-known object detectors: YOLOv2 and Faster R-CNN. We believe coarse annotations can be used in recognition systems based on the iris and periocular regions, given the much smaller engineering effort required to manually annotate the training images. We manually made coarse annotations of the iris and periocular regions ($\approx$ 122K images from the visible (VIS) spectrum and $\approx$ 38K images from near-infrared (NIR) spectrum). The iris annotations in the NIR databases were generated semi-automatically by first applying an iris segmentation CNN and then performing a manual inspection. These annotations were made for 11 well-known public databases (3 NIR and 8 VIS) designed for the iris-based recognition problem, and are publicly available to the research community[1]. Experimenting our proposal on these databases, we highlight two results. First, the Faster R-CNN + Feature Pyramid Network (FPN) model reported an Intersection over Union (IoU) higher than YOLOv2 (91.86% vs 85.30%). Second, the detection of the iris and periocular regions being performed simultaneously is as accurate as performed separately, but with a lower computational cost, i.e., two tasks were carried out at the cost of one.**

I. Introduction

In recent years, the interest in biometrics to automatically identify and/or verify a person's identity has greatly increased [1], [2]. Biometrics refers to the use of physiological and behavioral characteristics of humans for personal identification [3]. Such characteristics are particularly important since they cannot be changed, forgotten, lost or stolen, providing an unquestionable connection between the individual and the application that makes use of them [4].

Several characteristics of the human body can be used as biometrics such as fingerprints, face, ocular region components and voice, each with advantages and disadvantages. Among the aforementioned modalities, ocular biometric traits have received significant attention in the recent past [5]–[7] due to the fact that the ocular region is an important and interrelated human trait consisting of several parts, for example, the cornea, lens, optic nerve, retina, pupil, iris, and periocular region. In this direction, many authors proposed biometric systems based on iris, periocular, retina, and sclera regions, as they are considered potential biometric modalities [8], [9].

The iris appears as one of the main biological characteristics in security systems since it remains unchanged over time and its uniqueness level is high [10]. Furthermore, the identification using the iris region is non-invasive, that is, there is no need for physical contact to obtain and analyze an iris image [11]. However, after decades of research in personal identification, it has been observed that better results can be achieved by combining different biometric modalities [6], [12], [13]. A good example of it is the combination of iris and periocular-based biometrics [14], [15].

In this work, we compare the detection of the iris and periocular regions being performed separately and simultaneously using two well-known object detection networks: YOLOv2 [16] and Faster R-CNN [17]. Such deep models were chosen due to the fact that (i) promising results were recently obtained using them in other detection tasks [18]–[20]; and (ii) handcrafted features are easily affected by noise and might not be robust for unconstrained scenarios.

Typically, in biometric systems that use iris and/or periocular region images as input, the first step in which efforts should be applied is the detection of the Region of Interest (ROI) [21], as a poor detection would probably impair the effectiveness of the subsequent steps of the system [12], [22]. Recently, Zanlorensi et al. [23] showed that impressive iris recognition rates can be achieved when using deep representations having as system input the bounding boxes of the iris region, without the iris segmentation preprocessing. Also using deep representations and having as input a squared region (i.e., a bounding box), Luz et al. [24] achieved state-of-the-art results for periocular recognition. Such results, shorter execution times compared to single detection approaches (in which the iris and the periocular region are detected separately), and the promising results obtained in preliminary experiments support our motivation to detect both regions simultaneously.

The main contributions of this paper can be summarized as follows: (i) two new approaches for the simultaneous detection of the iris and periocular region; (ii) a comparative evaluation between detecting both regions simultaneously or separately in **eleven** publicly available databases; and (iii) for learning the models used in the experiments, coarse annotations (i.e., bounding boxes) were manually made for both iris and periocular regions of 122,738 images from 8 well-known visible (VIS) spectral databases. As stated by Cordts et al. [25], coarse annotations are intended to support research areas that exploit large volumes of data. We also automatically generated 38,851 bounding boxes using the iris segmentation

[1] All annotations made by us are publicly available at the following website: https://web.inf.ufpr.br/vri/databases/iris-periocular-coarse-annotations/.

approach proposed by Bezerra et al. [26] for 3 well-known near-infrared (NIR) spectral databases. We manually checked and corrected (if necessary) all annotations.

We chose the approach proposed in [26] due to the fact that it presented an error rate lower than 1.5% in the aforementioned NIR databases. However, despite the good results presented by that segmentation approach, the detection task is much less expensive in terms of both computational cost and data annotation. Regarding the 11 databases employed in our experiments, they were chosen because they are widely used in the biometric recognition literature [23], [27]–[29], which we plan to investigate in future works. It should be noted that, in many works in the literature, no more than three databases were used in the experiments [30]–[33].

In our experiments, the Faster R-CNN model yielded Intersection over Union (IoU) values higher than YOLOv2 (91.86% vs 85.30%) and the detection of the iris and periocular regions being performed simultaneously is as accurate as performed separately, but with a lower computational cost, i.e., two tasks were carried out at the cost of one. Regarding the use of coarse annotations, we believe they can be used in recognition systems based on the iris and periocular regions, given the much smaller engineering effort required to manually annotate the training images.

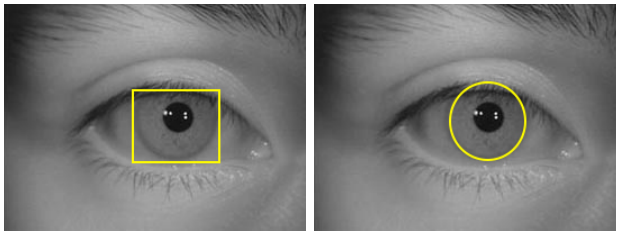
The remainder of this paper is organized as follows. We review related works in Section II. In Section III, our methodology is described. Section IV and Section V present, respectively, the experimental setup and the results obtained. Finally, conclusions and future work are discussed in Section VI.

II. Related Work

In this section, we discuss works related to iris and periocular region detection and conclude with final remarks.

A. Iris Detection

Regarding iris detection, the works in the literature commonly show the detected ROI using two different representations. Fig. 1a shows the use of a rectangular bounding box as the iris delimitation, while Fig. 1b shows an elliptical ROI detection using the outer iris boundary.



(a) Rectangular bounding box (b) Outer iris contour

Fig. 1. Samples of representation to iris ROI extraction.

Many works in the literature show the iris delimitation by using an elliptical contour around the outer edge of it. Daugman [21] pioneered this scenario by proposing an approach that makes the use of an integro-differential operator to detect the iris identifying the borders present in the images. This operator takes into account the circular shape of the iris to find its correct position by maximizing the partial derivative concerning the radius. In the experiments, the author employed a private database composed of 592 eye images captured in the NIR wavelength from 323 subjects.

Zhang & Ma [31] adopted a method that employs a momentum-based level set [34], [35] along with the Daugman's operator to locate the pupil boundary. Specifically, an initial contour of the iris is obtained with a momentum-based level set using the minimum average gray level. Then, the integro-differential operator is applied to perform the final detection, reducing the execution time and improving the results obtained in [21]. An accuracy rate of 98.53% was achieved on the CASIA-IrisV2 database [36]. Such improvement occurs because the initially detected contour is generally close to the actual inner boundary of the iris.

Alvarez-Betancourt & Garcia-Silvente [32], on the other hand, presented an iris location method based on the detection of circular boundaries through gradient analysis in points of interest of successive arcs, reaching an accuracy of 98% on the CASIA-IrisV3 database [37] with improvements in processing time. The quantified majority operator QMA-OWA, proposed in [38], was used to obtain a representative value for each successive arc. Then, the iris boundary is given by the arc with the most significant representative amount.

In the method proposed by Cui et al. [39], the eyelashes are removed as a first step using the dual-threshold method, which can be an advantage over other iris location approaches. Next, the facula is removed by using mathematical morphology. Finally, the accurate iris position is obtained through Hough-Transform and least-squares. Their method achieved 98% accuracy in the CASIA-IrisV3-Twins database [37].

Zhou et al. [33] presented a method for iris location in which the initial position of the iris is obtained by using the Vector Field Convolution (VFC) technique. This initial estimate makes pupil location much closer to the actual boundary instead of circle fitting, improving location accuracy and reducing computational cost. The final result is obtained using the algorithm proposed by Daugman, reducing the computational cost and improving the location accuracy since the pupil delineation is much closer to the actual boundary. An accuracy rate of 98.85% was reported on the CASIA-IrisV2 database.

Su et al. [40] proposed an iris location algorithm based on local property and iterative searching, achieving 98.08% accuracy on the CASIA-IrisV1 and CASIA-IrisV3 databases (i.e., they were combined in their experiments). In order to detect the ROI, the pupil area is extracted using iris regional attributes, and the inner edge of it is fitted by iterating, comparing and sorting the pupil edge points. The outer edge location is made by using an iterative searching method from the extracted pupil center and radius, with a shorter time in relation to the approaches available in the literature.

Chen & Ross [41] designed a multi-task Convolutional Neural Network (CNN)-based approach for joint iris and presentation attack detection. The experiments were performed on six publicly available databases, however, iris detection results were not reported as the main focus of their work is to identify presentation attacks.

Severo et al. [42] represented the iris as a rectangular bounding box. They fine-tuned the Fast-YOLOv2 model, which is much faster but less accurate than YOLOv2, in order to perform the ROI extraction, overcoming problems such as noise, eyelids, eyelashes and reflections. Six public databases were used to evaluate their method, which attained accuracy rates above 97% in all of them.

Wang et [43] recently introduced IrisParseNet, a network for iris detection that reached 89.40%, 85.39% and 85.07% IoU values in the CASIA-Iris-Distance, UBIRIS.v2 and MICHE-I databases, respectively. Their method simultaneously estimates the pupil center, the iris segmentation mask, and the iris inner/outer boundaries.

B. Periocular Region Detection

Park et al. [44] proposed one of the first biometric approaches based on the periocular region, featuring an eye region detector that uses face images detected by the Viola-Jones detector [45] as input and outputs the periocular region.

Similarly, Juefei-Xu & Savvides [46] also proposed a periocular region detection approach that employs as input a face image detected by the Viola-Jones detector. Nevertheless, the periocular region is identified using Active Shape Models (ASMs) that identify 79 facial landmarks, containing points relative to the eye region among them.

Mahalingam et al. [47] designed an eye detector that receives a face image and outputs the periocular region through Average of Synthetic Exact Filters (ASEF). All experiments were carried out on a private database composed of 1.2 million faces from 38 subjects. Le et al. [48], on the other hand, proposed a Local Eyebrow Active Shape Model (LE-ASM) to first detect the eyebrow region directly from a given face image and then to detect the periocular region using ASMs. The results obtained on this particular stage (i.e., periocular region detection) were not reported.

Proença et al. [49] proposed a Markov Random Field (MRF) method to segment the periocular region components and other elements around them (i.e., the iris, sclera, eyelashes, eyebrows, hair, skin and glasses). Their approach analyzes the image pixels and outputs the segmented region taking into account appearance and geometrical constraints and assuring that the system output is biologically plausible. The periocular region can be predicted by combining the outer limits of the sclera and the lower eyelashes.

C. Final Remarks

In most works, the accuracy was employed as the evaluation metric for iris and periocular region detection. However, the authors used different protocols to calculate the accuracy or do not specifically describe how the accuracy obtained by their approach was computed. Therefore, it is plausible to question how robust one method is compared to another.

While in the iris detection scenario a poor description of the evaluation metrics used has been made, in the periocular region detection scenario none of the studies found in the literature report the results achieved in this particular stage, probably due to the fact that the detection of the ROI was considered only as a preprocessing step in such works [44], [46]–[48].

Taking this information into consideration and also the fact that CNNs are not widely explored in the iris and periocular region detection domains, we propose to evaluate two well-known CNN object detectors (i.e., YOLOv2 and Faster R-CNN) in **eleven** coarsely annotated databases.

More specifically, the main objective of this work is to evaluate the simultaneous detection of the iris and periocular regions. The simultaneous detection approach is proposed taking into account the assumption that CNNs are able to understand the context present in the images, thus improving the results obtained by conventional single detection approaches.

III. Methodology

Currently, one of the most accurate ways to perform image classification, segmentation and object detection is using deep CNNs. Therefore, in this work, we propose the simultaneous detection of the iris and periocular regions using two object detection models: YOLOv2 [16] and Faster R-CNN [17]. It should be noted that (i) we trained both models from scratch; (ii) such models were chosen because promising results were obtained using them in other detection tasks [18]–[20].

Our hypothesis is that the proposed simultaneous detection approach is able to understand the context of the image and thereby improve detection results compared to single detection approaches in which the iris and the periocular region are detected separately. As baselines, we also adopted the YOLOv2 and Faster R-CNN models, but in two independent detection steps, i.e., one for the iris and one for the periocular region.

A. YOLOv2

Table I presents the YOLOv2 model, employed for detecting the iris and the periocular region. The architecture has 19 convolutional and 5 max-pooling layers. The convolutional layers, except for the last one, are divided into two groups: external and internal. The layers belonging to the external group use kernels of size 3×3, whereas the layers belonging to the internal group use kernels of size 1×1. Alternating 1×1 convolutional layers reduce the features space from preceding layers [50]. The convolutional blocks are composed of: convolution, batch normalization, and a Leaky Rectified Linear Unit (Leaky ReLU).

As this model does not have fully connected layers, it can receive images of any size as input. We adopted an input size of 416×416 pixels due to the good results achieved employing these dimensions in [16]. We also reduced the number of filters in the last convolutional layer to match our number of classes. The number of filters in that layer is given by

$$\textit{filters} = (C + 5) \times A \, , \tag{1}$$

where A is the number of anchor boxes (we use $A = 5$) used to predict bounding boxes and C is the number of classes, in our case either $C = 1$ or $C = 2$ to detect the iris and periocular regions separately or simultaneously, respectively. Thus, there are 30 filters in the last convolutional layer when the regions are detected separately and 35 when they are detected simultaneously.

The main difference between the YOLOv2 model proposed in [16] and the one used in this work is that we removed the route layers, i.e., layers that concatenate a list of previous layers together. In preliminary experiments, we observed that removing such layers did not negatively affect the results obtained in our tasks and also reduced the execution time.

TABLE I
THE YOLOv2 MODEL, MODIFIED FOR THE DETECTION OF THE IRIS AND THE PERIOCULAR REGION. THERE ARE 30 FILTERS IN THE LAST CONVOLUTIONAL LAYER WHEN THE REGIONS ARE DETECTED SEPARATELY AND 35 WHEN THEY ARE DETECTED SIMULTANEOUSLY.

#	Layer	Group	Filters	Size	Input	Output
0	conv	External	32	$3 \times 3/1$	$416 \times 416 \times 1/3$	$416 \times 416 \times 32$
1	max			$2 \times 2/2$	$416 \times 416 \times 32$	$208 \times 208 \times 32$
2	conv	External	64	$3 \times 3/1$	$208 \times 208 \times 32$	$208 \times 208 \times 64$
3	max			$2 \times 2/2$	$208 \times 208 \times 64$	$104 \times 104 \times 64$
4	conv	External	128	$3 \times 3/1$	$104 \times 104 \times 64$	$104 \times 104 \times 128$
5	conv	Internal	64	$1 \times 1/1$	$104 \times 104 \times 128$	$104 \times 104 \times 64$
6	conv	External	128	$3 \times 3/1$	$104 \times 104 \times 64$	$104 \times 104 \times 128$
7	max			$2 \times 2/2$	$104 \times 104 \times 128$	$52 \times 52 \times 128$
8	conv	External	256	$3 \times 3/1$	$52 \times 52 \times 128$	$52 \times 52 \times 256$
9	conv	Internal	128	$1 \times 1/1$	$52 \times 52 \times 256$	$52 \times 52 \times 128$
10	conv	External	256	$3 \times 3/1$	$52 \times 52 \times 128$	$52 \times 52 \times 256$
11	max			$2 \times 2/2$	$52 \times 52 \times 256$	$26 \times 26 \times 256$
12	conv	External	512	$3 \times 3/1$	$26 \times 26 \times 256$	$26 \times 26 \times 512$
13	conv	Internal	256	$1 \times 1/1$	$26 \times 26 \times 512$	$26 \times 26 \times 256$
14	conv	External	512	$3 \times 3/1$	$26 \times 26 \times 256$	$26 \times 26 \times 512$
15	conv	Internal	256	$1 \times 1/1$	$26 \times 26 \times 512$	$26 \times 26 \times 256$
16	conv	External	512	$3 \times 3/1$	$26 \times 26 \times 512$	$26 \times 26 \times 512$
17	max			$2 \times 2/2$	$26 \times 26 \times 512$	$13 \times 13 \times 512$
18	conv	External	1024	$3 \times 3/1$	$13 \times 13 \times 512$	$13 \times 13 \times 1024$
19	conv	Internal	512	$1 \times 1/1$	$13 \times 13 \times 1024$	$13 \times 13 \times 512$
20	conv	External	1024	$3 \times 3/1$	$13 \times 13 \times 512$	$13 \times 13 \times 1024$
21	conv	Internal	512	$1 \times 1/1$	$13 \times 13 \times 1024$	$13 \times 13 \times 512$
22	conv	External	1024	$3 \times 3/1$	$13 \times 13 \times 512$	$13 \times 13 \times 1024$
23	conv		$30/35$	$1 \times 1/1$	$13 \times 13 \times 1024$	$13 \times 13 \times 30/35$
24	detection					

B. Faster R-CNN + Feature Pyramid Network

We employ the Faster R-CNN model [17] combined with a Feature Pyramid Network (FPN) [51], as shown in Figure 3. Faster R-CNN is commonly composed of (i) a feature map extraction network; (ii) a region proposal network and (iii) a detection network. We replaced the standard CNN feature extraction module by an FPN, and thus multiple feature map layers are generated with better quality information than the regular implementation of Faster R-CNN.

C. Coarse Annotations

In this work, we use coarse annotations both to train and to evaluate our networks. As can be seen in Fig. 2, we define as a coarse annotation the region around the ROI so that the edges of the bounding box remain outside the limits of the fine annotations proposed by Severo et al. [42]. More specifically, the delimited region is larger than the one typically used in fine annotations, and the iris is not well-centered. Also, in some cases, the eyebrows were left out the ROI, as the images from some databases used in this work do not contain that region.

It is worth noting that the coarse annotations were made manually by two volunteers and that no strict rules of how annotations should be made were defined (besides simple instructions and the fact that were coarse and not fine annotations). Hence, there are random variations (in size, position, aspect ratio, etc.) among annotations of different images.

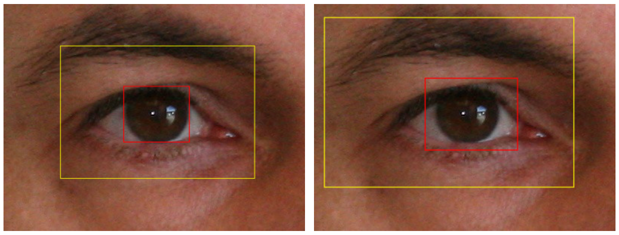

(a) Fine (b) Coarse

Fig. 2. Examples of fine and coarse annotations of both the iris (red) and the periocular region (yellow).

We believe coarse annotations can be used in recognition systems based on the iris and/or the periocular region, given the much smaller engineering effort required to manually annotate the training images. In other words, we conjecture that deep models for person identification may achieve promising results even when these regions are not perfectly segmented.

IV. Experimental Setup

In this section, we present the databases and also the evaluation protocol used in our experiments. The experiments were carried out on eleven databases, which are described in Section IV-A. Note that we trained/tested the networks on each dataset separately. All experiments were performed on a computer with an Intel® Core™ i7-7700 4.20GHz CPU, 16 GB of RAM and two NVIDIA Titan Xp GPUs.

A. Databases

We employed the following public databases: CASIA-Iris-Interval [37], CASIA-Iris-Lamp [37], CASIA-Iris-Thousand [37], Cross-Eyed-VIS [52], CSIP [53], MICHE-I [54], MobBIO [55], NICE-II [56], PolyU-VIS [57], UBIRIS.v2 [56] and VISOB [58]. An overview of the important features of all databases used in this work can be seen in Table II. These databases were chosen because they are widely used in the biometric recognition literature [23], [27]–[29], which we plan to investigate in future works.

CASIA-Iris-Interval: the iris images of this database were captured with a close-up iris camera developed by the authors themselves. The database consists of 2,639 images from 249 subjects and 395 classes, with a resolution of 320×280 pixels, obtained in two sections.

CASIA-Iris-Lamp: the images were collected using a non-fixed sensor and, thus, the individuals collected the iris image with the sensor in their own hands. While capturing the images, a lamp was turned on and off in order to produce more intraclass variations due to pupil contraction and expansion, creating a nonlinear deformation. A total of 16,212 images with a resolution of 640×480 pixels from 411 subjects and 819 classes were collected in a single section.

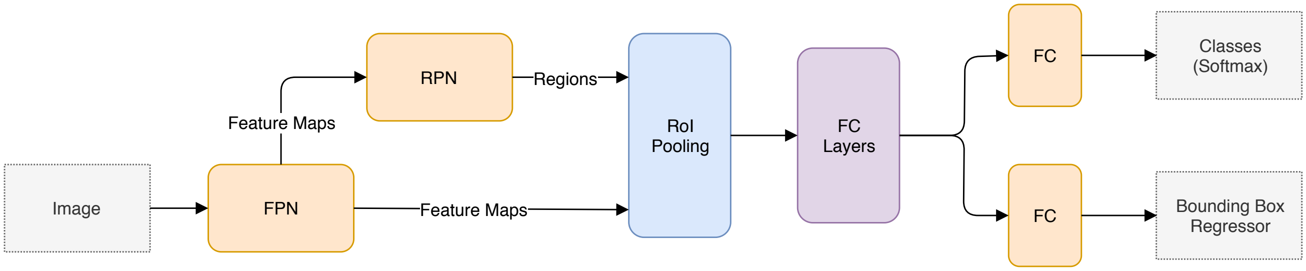


Fig. 3. Faster R-CNN + FPN architecture overview.

TABLE II
OVERVIEW OF THE IMPORTANT FEATURES OF THE DATABASES USED IN THIS WORK. ALL OF THESE ARE A SUBSET OF THE ORIGINAL DATABASE.

Database	Year	Images	Subjects	Resolution	Wavelength
CASIA-Iris-Interval [37]	2010	2,639	249	320 × 280	NIR
CASIA-Iris-Lamp [37]	2010	16,212	411	640 × 480	NIR
CASIA-Iris-Thousand [37]	2010	20,000	1,000	640 × 480	NIR
Cross-Eyed-VIS [52]	2016	1,920	120	400 × 300	VIS
CSIP * [53]	2015	2,004	50	Various	VIS
MICHE-I * [54]	2015	3,191	92	Various	VIS
MobBIO [55]	2014	1,206	105	300 × 200	VIS
NICE-II [56]	2010	2,000	n/a	400 × 300	VIS
PolyU-VIS [57]	2017	6,270	209	640 × 480	VIS
UBIRIS.v2 [56]	2010	11,101	261	400 × 300	VIS
VISOB* [58]	2016	95,046	550	Various	VIS
NIR		38,851			
VIS		122,738			
Total		161,590			

* Cross-sensor databases

CASIA-Iris-Thousand: this database contains 20,000 iris images from 1000 subjects with a resolution of 640 × 480 pixels, which were collected in a single section using an IKEMB-100 camera.

Cross-Eyed-VIS: this database subset is composed of VIS images. Eight images of each eye were captured from 120 subjects, totaling 1,920 images. The images have dimensions of 400 × 300 pixels. All images were obtained at a distance of 1.5 meters, in an uncontrolled indoor environment, with a wide variation of ethnicity, eye colors, and lighting conditions.

CSIP: this database has images acquired with four different mobile devices: *Sony Ericsson Xperia Arc S* (rear 3,264 × 2,448), *iPhone* 4 (front 640 × 480, rear 2,592 × 1,936), *THL W*200 (front 2,592 × 1,936, rear 3,264 × 2,448), and *Huawei U*8510 (front 640 × 480, rear 2,048 × 1,536). The database has 2,004 images from 50 subjects.

MICHE-I: this database contains 3,732 images from 92 subjects acquired by mobile devices in visible light. In order to simulate a real application, the iris images were obtained by the users themselves, indoors and outdoors, with and without glasses. Images of only one eye of each individual were captured. The mobile devices used and their respective resolutions are the following: *iPhone* 5 (1,536 × 2,048), *Samsung Galaxy S*4 (2,322 × 4,128) and *Samsung Galaxy Tablet II* (640 × 480).

MobBIO: this database has face, iris, and voice biometric data belonging to 105 subjects. The data was acquired with the mobile device *Asus Transformer Pad (TF*300*T)*. The iris images were obtained in two different lighting conditions, with varying eye orientations and occlusion levels. For each subject, 16 images (8 of each eye) were captured.

NICE-II: this database, a subset of UBIRIS.v2, contains 2,000 images with a resolution of 400 × 300 pixels and was employed in the NICE.II contest. The number of subjects of this set was not directly specified.

PolyU-VIS: this database has 6,270 iris images with a resolution of 640 × 480 pixels, with 15 images of each eye from 209 subjects obtained in the visible spectrum [57].

UBIRIS.v2: this database contains 11,101 RGB images captured with a Canon EOS 5D camera and resolution of 400 × 300 pixels, from 261 subjects (i.e., 522 irises) [56].

VISOB: front cameras of three mobile devices were used to obtain the images of this database, such as the iPhone 5S at 720p resolution, Samsung Note 4 at 1080p resolution and Oppo N1 also at 1080p resolution. The images were captured in 2 sessions for each of the 2 visits, which occurred between 2 and 4 weeks, totaling 158,136 images from 550 subjects.

B. Evaluation Protocol

The evaluation of an automatic detection approach is performed in a pixel-to-pixel comparison between the ground truth and the predicted bounding boxes. Therefore, we use the mean F-score, IoU and mean Average Precision (mAP) evaluation metrics. Following Severo et al. [42], to first compute the precision and recall metrics and then the F-score, we consider as correct the bounding boxes detected with an IoU value above 0.5 with the ground truth. This bounding box evaluation, defined in the PASCAL VOC Challenge [59], is interesting since it penalizes both over- and under-estimated objects.

It is worth noting that we use coarse annotations as the ground truth, as the databases do not provide fine annotations of the position of the iris and periocular regions on each image. In this sense, instead of evaluating the predicted bounding boxes in relation to the exact location of the iris/periocular region, we evaluated how close to the ground truth it is.

In order to perform a fair evaluation and comparison of the proposed approaches, we divided each database into three subsets, being 40% of the images for training, 40% for testing and 20% for validation. We adopt this protocol (i.e., with a larger test set) to provide more samples for analysis of statistical significance. Also, in the statistical direction, we perform the Wilcoxon signed-rank test [60] to verify if there is a statistical difference between the detection approaches.

TABLE III

DETECTION RESULTS. THE SINGLE AND MULTI COLUMNS PRESENT THE RESULTS OBTAINED WHEN DETECTING THE IRIS AND PERIOCULAR REGIONS SEPARATELY AND SIMULTANEOUSLY, RESPECTIVELY. THE VALUES IN BOLD REPRESENT THE HIGHEST IOU VALUES OBTAINED, WHILE THE HIGHLIGHTED RESULTS INDICATE THE CASES IN WHICH THERE IS NO STATISTICAL DIFFERENCE ACCORDING TO THE WILCOXON STATISTICAL TESTS.

	F-score				IoU (%)				mAP (%)			
Database	YOLOv2		Faster R-CNN		YOLOv2		Faster R-CNN		YOLOv2		Faster R-CNN	
	Multi	Single	Multi	Single	Multi	Single	Multi	Single	Multi	Single	Multi	Single
						Iris						
CASIA-Iris-Interval	0.90	0.92	0.97	0.96	82.81	86.20	**94.77**	93.98	100.00	100.00	100.00	93.98
CASIA-Iris-Lamp	0.96	0.96	0.98	0.95	92.38	93.06	96.08	**97.31**	99.98	99.98	99.73	97.31
CASIA-Iris-Thousand	0.97	0.98	0.98	0.98	95.71	94.39	**97.72**	97.58	99.96	99.97	99.65	97.58
Cross-Eyed-VIS	0.92	0.92	0.94	0.94	85.79	86.45	90.39	**90.44**	100.00	100.00	100.00	90.44
CSIP	0.92	0.73	0.95	0.95	87.97	58.12	**91.61**	91.55	98.68	98.69	100.00	91.55
MICHE-I	0.88	0.83	0.92	0.92	80.32	72.07	86.27	**86.48**	97.39	94.32	100.00	92.48
MobBIO	0.95	0.95	0.96	0.96	91.52	91.40	**94.14**	93.79	100.00	100.00	100.00	93.79
NICE-II	0.90	0.91	0.93	0.82	83.39	84.83	**88.41**	78.20	98.92	99.32	99.32	78.20
PolyU-VIS	0.91	0.86	0.94	0.94	**93.81**	76.32	89.12	89.31	99.74	93.79	100.00	89.31
UBIRIS.v2	0.89	0.89	0.91	0.91	81.16	81.75	85.16	**85.26**	99.35	99.00	100.00	85.29
VISOB	0.91	0.89	0.96	0.96	85.04	81.32	**93.09**	92.80	99.53	99.34	99.90	92.80
						Periocular Region						
CASIA-Iris-Interval	0.96	0.98	0.98	0.98	92.65	96.19	**97.80**	96.79	98.62	100.00	100.00	97.80
CASIA-Iris-Lamp	0.98	0.97	0.99	0.98	97.15	96.02	**98.08**	97.71	99.95	99.95	99.97	97.70
CASIA-Iris-Thousand	0.97	0.98	0.99	0.99	95.92	96.44	**98.19**	98.19	99.89	99.94	99.97	98.18
Cross-Eyed-VIS	0.92	0.92	0.96	0.96	86.86	86.89	**92.74**	92.56	97.84	99.66	100.00	92.56
CSIP	0.95	0.95	0.87	0.96	91.61	91.76	84.97	**92.96**	99.83	100.00	83.61	92.96
MICHE-I	0.85	0.85	0.90	0.90	75.88	74.97	**83.66**	83.51	93.82	96.33	98.77	93.51
MobBIO	0.96	0.96	0.97	0.97	94.21	94.09	**95.50**	94.83	100.00	100.00	100.00	94.83
NICE-II	0.88	0.90	0.92	0.92	80.52	82.44	**86.91**	86.66	97.23	99.55	99.76	86.66
PolyU-VIS	0.96	0.87	0.98	0.98	93.57	77.95	**96.74**	96.41	99.48	99.56	100.00	96.41
UBIRIS.v2	0.87	0.88	0.91	0.91	78.98	80.03	85.19	**85.44**	83.12	98.35	99.64	85.44
VISOB	0.93	0.94	0.97	0.98	87.17	89.11	96.08	**96.35**	95.64	99.98	99.83	96.35

V. RESULTS

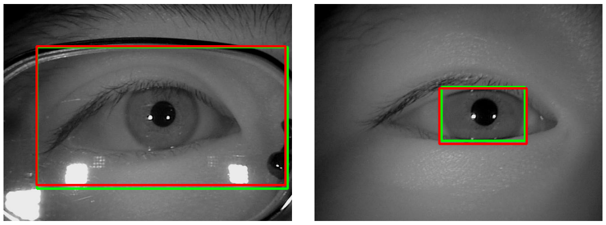
The experiments were carried out using the protocol presented in Section IV-B. To compare the proposed approaches, we report the F-score values in order to analyze the trade-off between precision and recall measures, however, we focus on the IoU metric since we want to assess how close are the predicted bounding boxes compared to the ground truth.

When analyzing the results regarding ***iris*** detection (see top of Table III), in 10 of 11 experiments the highest mean IoU value was achieved using Faster R-CNN. In general, the best results were obtained when simultaneously detecting the iris and the periocular region. The exceptions are in the CASIA-Iris-Lamp, Cross-Eyed-VIS, MICHE-I, and UBIRIS.v2 databases, where detecting both regions separately performed better, probably due to the fact that there are not many variations in iris and periocular region arrangement in the images of these databases. However, as the difference in the results obtained with both approaches is very small, we applied the Wilcoxon signed-rank test and observed that there is no statistical difference between detecting the iris and the periocular region simultaneously or separately in the CASIA-Iris-Lamp, Cross-Eyed-VIS, CSIP and MobBIO databases. In this way, in Table III, we highlighted (light gray) the results obtained in these databases.

Similar behavior occurred in the detection of the ***periocular*** region, however, in this case, all the best results were attained employing the Faster R-CNN model. In this scenario, the detection results using the single-class detection approach CSIP, UBIRIS.v2 and VISOB databases presented the best values. Similar to the results on iris detection, the difference between the IoU values attained between the approaches is close and there is no statistical difference in the CASIA-Iris-Thousand, Cross-Eyed-VIS and NICE-II databases and that result was also highlighted in Table III.

We emphasize that most of the best results were obtained using the Faster R-CNN + FPN approach, which we believe to be justified by the fact that FPNs perform a better feature map extraction compared to other approaches [51].

It should be noted that the IoU values obtained were higher than 95% for both iris and periocular region detection in the databases where the images were captured using a NIR sensor. These results were achieved by using the Faster R-CNN simultaneous detection approach, and the better detected iris and periocular region can be seen in Figure 4.



(a) Periocular Region

(b) Iris Detection

Fig. 4. Best iris and periocular region detection performed by the Faster R-CNN simultaneous detection approach. The green bounding boxes represent the coarse annotations, while the red ones represent the detected regions.

Despite the good results it is necessary observe that in the databases in which the images were captured using more than one sensor, that there was no preprocessing of the image (i.e., MICHE-I and CSIP) or that composed with lower quality images (i.e., UBIRIS.v2 and NICE-II) we obtained results with IoU values lower than 90% when detecting both the iris and periocular regions simultaneously. By analyzing these images, we can understand what made the results obtained by the approaches on these databases below than 90% of IoU: i) the use of eyeglasses; ii) the presence of more than one eye;

VI. Conclusions

In this work, we compared the detection of the iris and the periocular region being performed separately or simultaneously using two well-known object detectors, observing a better performance of the Faster R-CNN + FPN approach.

The detection of both regions being performed simultaneously produced better results in most databases, for both the iris and the periocular region. This leads us to believe that using this approach gives the neural network a certain understanding of the context present in the image.

We also coarsely labeled 161,590 images for iris and periocular region detection. These annotations are publicly available to the research community, assisting the development and evaluation of new detection approaches as well as the fair comparison among published works.

There is still room for improvements in the simultaneous detection of iris and periocular region. As future work, we intend to (i) design new and better network architectures; (ii) design a general and independent sensor approach, where the image sensor is first classified and then the iris and the periocular region are simultaneously detected with a specific approach; (iii) compare the proposed approach with methods applied in other domains; (iv) create a context-aware object-detection architecture; and (v) design a cascade detection approach for iris and periocular region detection.

Acknowledgements

This work was supported by the National Council for Scientific and Technological Development (CNPq) (grant numbers 428333/2016-8 and 313423/2017-2) and the Coordination for the Improvement of Higher Education Personnel (CAPES) (Social Demand Program). The Titan Xp GPUs used for this research were donated by the NVIDIA Corporation.

References

[1] A. Abdelwhab and S. Viriri, "A survey on soft biometrics for human identification," in *Machine Learning and Biometrics*, 2018.

[2] K. W. Bowyer and M. J. Burge, *Handbook of iris recognition*, 2016.

[3] A. Das, U. Pal, M. A. F. Ballester, and M. Blumenstein, "Sclera recognition using dense-SIFT," in *International Conference on Intelligent Systems Design and Applications*, Dec 2013, pp. 74–79.

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

[5] A. Das, U. Pal, M. A. F. Ballester, and M. Blumenstein, "Multi-angle based lively sclera biometrics at a distance," in *IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM)*, Dec 2014, pp. 22–29.

[6] I. Nigam, M. Vatsa, and R. Singh, "Ocular biometrics: A survey of modalities and fusion approaches," *Information Fusion*, vol. 26, pp. 1–35, 2015.

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

[8] R. Hill, "Apparatus and method for identifying individuals through their retinal vasculature patterns," 1978, US Patent 4,109,237.

[9] U. Park, A. Ross, and A. K. Jain, "Periocular biometrics in the visible spectrum: A feasibility study," in *IEEE International Conference on Biometrics: Theory, Applications, and Systems*, Sep. 2009, pp. 1–6.

[10] Yong Zhu, Tieniu Tan, and Yunhong Wang, "Biometric personal identification based on iris patterns," in *International Conference on Pattern Recognition (ICPR)*, vol. 2, Sep. 2000, pp. 801–804.

[11] A. K. Jain, R. Bolle, and S. Pankanti, *Biometrics, Personal Identification in Networked Society*. Kluwer Academic Publishers, 1998.

[12] C. Tan and A. Kumar, "Human identification from at-a-distance images by simultaneously exploiting iris and periocular features," in *International Conference on Pattern Recognition*, Nov 2012, pp. 553–556.

[13] K. I. Chang, K. W. Bowyer, P. J. Flynn, and X. Chen, "Multi-biometrics using facial appearance, shape and temperature," in *International Conference on Automatic Face and Gesture Recognition*, 2004, pp. 43–48.

[14] L. Xiao, Z. Sun, and T. Tan, "Fusion of iris and periocular biometrics for cross-sensor identification," in *Biometric Recognition*, 2012, pp. 202–209.

[15] M. D. Marsico, M. Nappi, and H. Proença, "Results from MICHE II – Mobile Iris CHallenge Evaluation II," *Pattern Recognition Letters*, vol. 91, pp. 3–10, 2017.

[16] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *IEEE Conference on Computer Vision and Pattern Recognition (CPVR)*, 2017, pp. 6517–6525.

[17] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, June 2017.

[18] Y. Ding, Q. Tao, L. Wang, D. Li, and M. Zhang, "Image-based localisation using shared-information double stream hourglass networks," *Electronics Letters*, vol. 54, no. 8, pp. 496–498, 2018.

[19] K. E. Ko and K. B. Sim, "Real-time object entity detection system for smart surveillance application," *Electronics Letters*, vol. 53, no. 19, pp. 1304–1306, 2017.

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

[21] J. G. 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, pp. 1148–1161, Nov 1993.

[22] J. Daugman, "How iris recognition works," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, pp. 21–30, 2004.

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

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

[25] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 3213–3223.

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

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

[28] N. Aginako, J. M. Martínez-Otzeta, I. Rodriguez, E. Lazkano, and B. Sierra, "Machine learning approach to dissimilarity computation: Iris matching," in *International Conference on Pattern Recognition (ICPR)*, Dec 2016, pp. 170–175.

[29] A. Deshpande, S. Dubey, H. Shaligram, A. Potnis, and S. Chavan, "Iris recognition system using block based approach with DWT and DCT," in *IEEE Anual India Conference*, Dec 2014, pp. 1–5.

[30] J. L. G. Rodríguez and Y. D. Rubio, "A new method for iris pupil contour delimitation and its application in iris texture parameter estimation," in *Progress in Pattern Recognition, Image Analysis and Applications*. Springer Berlin Heidelberg, 2005, pp. 631–641.

[31] W. Zhang and Y. D. Ma, "A new approach for iris localization based on an improved level set method," in *International Computer Conference on Wavelet Actiev Media Technology and Information Processing*, 2014, pp. 309–312.

[32] Y. Alvarez-Betancourt and M. Garcia-Silvente, "A fast iris location based on aggregating gradient approximation using QMA-OWA operator," in *International Conference on Fuzzy Systems*, July 2010, pp. 1–8.

[33] L. Zhou, Y. Ma, J. Lian, and Z. Wang, "A new effective algorithm for iris location," in *IEEE ROBIO*, 2013, pp. 1790–1795.

[34] G. Läthén, T. Andersson, R. Lenz, and M. Borga, "Momentum based optimization methods for level set segmentation," in *Scale Space and Variational Methods in Computer Vision*. Springer Berlin Heidelberg, 2009, pp. 124–136.

[35] Zhejin Wang, Y. Feng, and Qinqin Tao, "Momentum based level set segmentation for complex phase change thermography sequence," in *International Conference on Computer Application and System Modeling*, vol. 12, Oct 2010, pp. 257–260.

[36] CASIA. (2004) Casia version 2 database. [Online]. Available: http://biometrics.idealtest.org/dbDetailForUser.do?id=2

[37] ——, "Casia version 4 database," 2010. [Online]. Available: http://biometrics.idealtest.org/dbDetailForUser.do?id=4

[38] J. Peláez and J. Doña, "A majority model in group decision making using QMA–OWA operators," *International Journal of Intelligent Systems*, vol. 21, no. 2, pp. 193–208, 2006.

[39] ZhuYu and Wang Cui, "A rapid iris location algorithm based on embeded," in *International Conference on Computer Science and Information Processing (CSIP)*, Aug 2012, pp. 233–236.

[40] L. Su, J. Wu, Q. Li, and Z. Liu, "Iris location based on regional property and iterative searching," in *IEEE International Conference on Mechatronics and Automation (ICMA)*, Aug 2017, pp. 1064–1068.

[41] C. Chen and A. Ross, "A multi-task convolutional neural network for joint iris detection and presentation attack detection," in *IEEE Winter Applications of Computer Vision Workshops*, March 2018, pp. 44–51.

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

[43] C. Wang, Y. Zhu, Y. Liu, R. He, and Z. Sun, "Joint iris segmentation and localization using deep multi-task learning framework," *CoRR*, 2019. [Online]. Available: http://arxiv.org/abs/1901.11195

[44] U. Park, R. R. Jillela, A. Ross, and A. K. Jain, "Periocular biometrics in the visible spectrum," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 1, pp. 96–106, 2011.

[45] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," *International Journal of Computer Vision*, no. 2, pp. 137–154, 2004.

[46] F. Juefei-Xu and M. Savvides, "Unconstrained periocular biometric acquisition and recognition using COTS PTZ camera for uncooperative and non-cooperative subjects," in *IEEE Workshop on the Applications of Computer Vision (WACV)*, Jan 2012, pp. 201–208.

[47] G. Mahalingam, K. Ricanek, and A. M. Albert, "Investigating the periocular-based face recognition across gender transformation," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2180–2192, 2014.

[48] T. H. N. Le, U. Prabhu, and M. Savvides, "A novel eyebrow segmentation and eyebrow shape-based identification," in *IEEE International Joint Conference on Biometrics (IJCB)*, 2014, pp. 1–8.

[49] H. Proença, J. C. Neves, and G. Santos, "Segmenting the periocular region using a hierarchical graphical model fed by texture/shape information and geometrical constraints," in *IEEE International Joint Conference on Biometrics (IJCB)*, 2014, pp. 1–7.

[50] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CPVR)*, 2016, pp. 779–788.

[51] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, July 2017, pp. 936–944.

[52] A. Sequeira *et al.*, "Cross-eyed - cross-spectral iris/periocular recognition database and competition," in *International Conference of the Biometrics Special Interest Group (BIOSIG)*, Sep. 2016, pp. 1–5.

[53] G. Santos, E. Grancho, M. V. Bernardo, and P. T. Fiadeiro, "Fusing iris and periocular information for cross-sensor recognition," *Pattern Recognition Letters*, vol. 57, pp. 52–59, 2015.

[54] M. De Marsico, M. Nappi, D. Riccio, and H. Wechsler, "Mobile Iris Challenge Evaluation (MICHE)-I, biometric iris dataset and protocols," *Pattern Recognition Letters*, vol. 57, pp. 17–23, 2015.

[55] A. F. Sequeira, J. C. Monteiro, A. Rebelo, and H. P. Oliveira, "MobBIO: A Multimodal Database Captured with a Portable Handheld Device," in *International Conference on Computer Vision Theory and Applications (VISAPP)*, vol. 3, 2014, pp. 133–139.

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

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

[58] A. Rattani, R. Derakhshani, S. K. Saripalle, and V. Gottemukkula, "ICIP 2016 competition on mobile ocular biometric recognition," in *IEEE International Conference on Image Processing*, 2016, pp. 320–324.

[59] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (VOC) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, Jun 2010.

[60] F. Wilcoxon, S. Katti, and R. A. Wilcox, "Critical values and probability levels for the Wilcoxon rank sum test and the Wilcoxon signed rank test," *Selected tables in mathematical statistics*, vol. 1, pp. 171–259, 1970.