Research Article

Impact of Occlusion Masks on Gender Classification from Iris Texture

Claudio Yáñez [iD],[1] **Juan E. Tapia** [iD],[2] **Claudio A. Perez** [iD],[1] and **Christoph Busch** [iD][2]

[1]*Department of Electrical Engineering, and Advanced Mining Technology Center, IMPACT, Center of Interventional Medicine for Precision and Advanced Cellular Therapy, Universidad de Chile, Santiago, Chile*
[2]*da/sec-Biometrics and Internet Security Research Group, Hochschule Darmstadt, Darmstadt University of Applied Sciences, Darmstadt, Germany*

Correspondence should be addressed to Juan E. Tapia; juan.tapia-farias@h-da.de

Gender classification on normalized iris images has been previously attempted with varying degrees of success. In these previous studies, it has been shown that occlusion masks may introduce gender information; occlusion masks are used in iris recognition to remove non-iris elements. When, the goal is to classify the gender using exclusively the iris texture, the presence of gender information in the masks may result in apparently higher accuracy, thereby not reflecting the actual gender information present in the iris. However, no measures have been taken to eliminate this information while preserving as much iris information as possible. We propose a novel method to assess the gender information present in the iris more accurately by eliminating gender information in the masks. This consists of pairing iris with similar masks and different gender, generating a paired mask using the OR operator, and applying this mask to the iris. Additionally, we manually fix iris segmentation errors to study their impact on the gender classification. Our results show that occlusion masks can account for 6.92% of the gender classification accuracy on average. Therefore, works aiming to perform gender classification using the iris texture from normalized iris images should eliminate this correlation.

1. Introduction

Normalized iris images are commonly used for subject identification or gender classification [1]. To achieve this, a periocular image is initially obtained. The iris in this image is then segmented and normalized [2]. As part of the segmentation and normalization processes, a mask that occludes non-iris regions is obtained to eliminate artifacts such as light reflections, eyelashes, and eyelids [2, 3]. This occlusion mask is important because it prevents non-iris information from interfering with the identification processes that follow the study by Li and Savvides [3]. After normalizing the iris image, an encoded version is generated, which is used alongside the occlusion mask to identify the subject. This process is summarized in Figure 1.

As a source of biometric information, the iris pattern offers multiple benefits [2]. It has enormous variability, which facilitates recognition [2]. It is also well-protected from the environment and stable over time [2]. Gender classification from iris benefits from these qualities and can

provide complementary information to recognition [4]. For instance, gender could be used as an additional trait for identity confirmation, preventing potential false matches [4]. It could also be used to speed up verification by searching only among the subjects with matching gender. Another potential application is to label and add demographic information to a previously unlabeled database.

Gender classification from iris has been addressed in several publications [5–8], with different degrees of success [4]. Studies have revealed that most gender information is not actually contained in the iris, but outside of it [4, 9–12]. However, performing gender classification, using exclusively the iris texture poses an interesting challenge, considering that differences in iris texture across genders have been reported in the medical literature [13]. By isolating the iris texture for gender classification, we aim to discern the extent to which gender cues are localized within this specific region. Furthermore, by using only the iris texture, we can benefit from the robustness and security associated with iris biometrics.
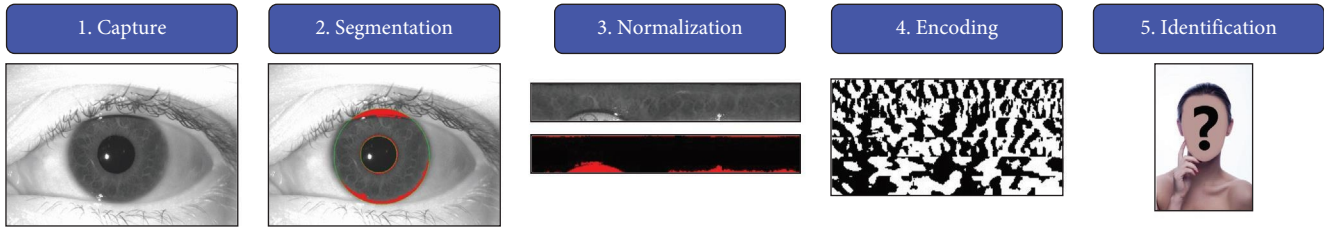
| 1. Capture | 2. Segmentation | 3. Normalization | 4. Encoding | 5. Identification |

FIGURE 1: Stages of iris processing and recognition.

(a)   (b)
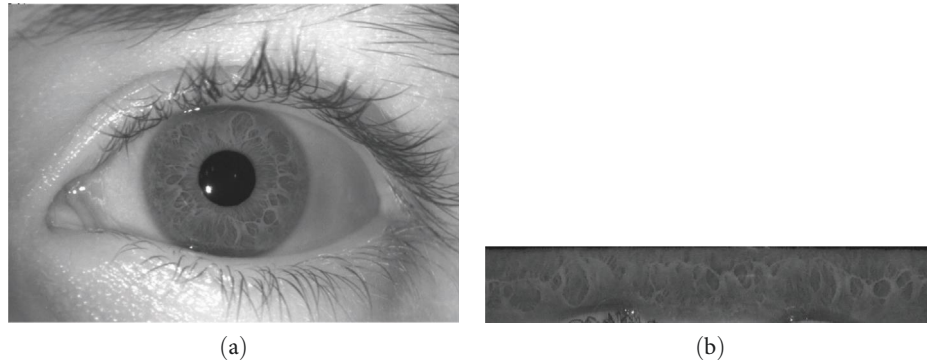
FIGURE 2: Periocular image (a) and normalized iris image (b).

It has been reported that the use of occlusion masks during gender classification from iris may introduce gender information, e.g., because of the use of cosmetics [14]. If the goal is to classify gender using exclusively the iris texture, steps must be taken to remove this additional information. This way the results will more closely reflect the gender information in the iris.

Another possible source of additional information is from automatically generated masks, which may not properly cover non-iris regions. These regions may contain gender cues, such as presence of makeup, eyelashes length, or eyelid texture, which could influence gender classification results. In case these non-iris regions are present, the results may not be reflecting gender information just from the iris, but also from the other regions [4, 14].

In this paper, we propose a novel method to eliminate gender information present in the masks by generating mask pairs during model training. This is performed by pairing every male mask with a similar female mask, and then generating a new mask by applying the union of the original two masks. By doing this, mask distributions become equal for both genders, thus neutralizing possible gender information from the masks. Furthermore, by pairing similar masks, the amount of iris information that is lost when generating the new masks is reduced.

Additionally, we created another set of masks where we manually corrected the automatically generated masks. This allows us to study the effect of errors in automatically generated masks. Both factors, mask pairs and mask manual corrections, provide new insights into the effect of masks in gender classification from iris.

The contributions of this paper are the following:

(i) A method, called mask pairing, for eliminating gender information in the occlusion masks, which allows for studying gender information exclusively in the iris texture.

(ii) A study of the impact of occlusion masks in gender classification.

(iii) An open-source toolbox for aiding manual mask correction (https://github.com/Nosferath/fixMasks).

The rest of this paper is organized as follows: Section 2 describes previous related work. Section 3 describes the employed methodologies and the datasets used in this paper. Section 4 shows and discusses the results of our experiments. Finally, Section 5 discusses our conclusions and proposes future work.

2. Related Work

Gender classification using iris images has been addressed in several studies following two main approaches [4]: classification using periocular iris images, and classification using normalized iris images. Classification using periocular iris images usually yields results close to and above 80% accuracy and has been repeated successfully over time [4, 9–12, 15–27]. Using periocular images for gender classification benefits from additional gender cues that are not present in the iris. For instance, Figure 2 shows a periocular image and a normalized iris image. The periocular image contains information that is additional to the iris, including skin texture, lacrimal caruncle, lacrimal puncta, sclera, pupil, eyebrows, eyelashes, and orbital bone structure cues [4, 9]. In contrast, the normalized iris image contains only the iris texture, and small portions of the eyelids, eyelashes, sclera, and pupil.

TABLE 1: Summary of works on gender classification from iris.

Citation	Year	Spectrum	Image type	Occlusion masks	Best accuracy (%)
[1]	2015	NIR	Normalized	✓	91.33
[35]	2015	NIR	Normalized	✗	89.74
[30]	2016	NIR	Normalized	✓	91.00
[9]	2016	NIR	Both	✗	69 Normalized, 85.7 periocular
[15]	2017	NIR	Periocular	✗	83.17
[10]	2017	NIR/VIS	Periocular	✗	89.59
[16]	2017	VIS	Periocular	✗	90.20
[31]	2017	NIR	Normalized	✓	84.66
[14]	2017	NIR	Both	✓	66 Normalized, 80 periocular
[17]	2018	VIS	Periocular	✗	90.00
[18]	2018	NIR	Periocular	✗	87.26
[11]	2018	NIR	Periocular	✗	85.93
[12]	2018	NIR	Periocular	✗	85.90
[19]	2019	NIR	Periocular	✗	85.40
[32]	2019	NIR	Normalized	✓	95.45
[4]	2019	NIR	Both	✓	63.40 Normalized, 80.80 periocular
[20]	2019	NIR	Periocular	✗	94.63
[21]	2019	NIR/VIS	Periocular	✗	86.89
[22]	2019	NIR/VIS	Periocular	✗	91.90
[23]	2019	NIR	Periocular	✗	91.90
[36]	2019	NIR	Normalized	✓	94.66
[24]	2019	VIS	Periocular	✗	90.15
[25]	2019	NIR	Both	✗	66.67 Normalized, 93.34 face
[26]	2020	VIS	Periocular	✗	92.00
[27]	2020	NIR/VIS	Periocular	✗	81.59
[37]	2020	NIR	Normalized	✗	96.00
[38]	2021	VIS	Periocular	✗	75.10
[39]	2021	NIR/VIS	Periocular	✗	75.72
[40]	2023	NIR	Both	✗	98.92 Normalized, 73.96 periocular
[41]	2023	VIS	Periocular	✗	98.71

Works that do not use the NIR spectrum have not been thoroughly included. Image type "Both" means experiments were performed with normalized and periocular images.

Studies that perform gender classification using normalized iris images have reported mixed results. Some [1, 28–, 32] report accuracies comparable to those obtained using periocular images, while others [4, 9, 14] report results closer to 60%. Explanations for these discrepancies have been briefly discussed in the past [4, 14], including possible non–subject-disjoint partitions, presence of non-iris information, or performing only one data partition.

Occlusion masks are used in iris recognition in order to remove non-iris elements [2, 33]. When the iris image is processed, an occlusion mask is generated during the iris segmentation stage; this can be done using various techniques depending on the algorithm [33, 34]. The occlusion mask is normalized alongside its corresponding iris. To perform iris recognition, a binary code is extracted from the normalized iris by applying Gabor filters to the iris and encoding the result. When two iris images are compared during the recognition process, both of their binary codes and occlusion masks are used. The masks are used to indicate which pixels should be accounted during comparison and which should be ignored. This improves recognition rates by preventing non-iris elements from influencing iris comparisons [2, 3].

Not all publications on gender classification have included occlusion masks (Table 1). Those are usually because of the software used to normalize the iris images. As a result, all the publications that we report using occlusion masks were on the normalized iris images. They have used either IrisBEE [42, 43] or OSIRIS [33, 44, 45] software, both of which were originally intended for iris recognition.

Earlier works on gender classification that used occlusion masks [1, 30] used the IrisBEE software to segment and normalize iris images. Later versions of IrisBEE added an occlusion mask that is based not only on iris segmentation, but also on so-called "fragile bits", bits in the iris code that are more prone to change due to imaging noise or occlusions [46]. In [1, 30], the location of fragile bits was used to perform gender classification. The occlusion mask was applied before feature extraction is performed, which hides the intensity value of these pixels while preserving their location. Neither of these publications comment on the impact of the occlusion masks on gender classification, although Tapia et al. [30] mentions

that the segmentation algorithm does not effectively mask out eyelashes when there is presence of makeup.

Kuehlkamp et al. [14] showed how makeup can cause errors in segmentation, which directly impacts occlusion masks. As a result, they demonstrated that occlusion masks contain information correlated with gender. This was done by performing gender classification using only the occlusion masks, which yielded up to 65% of accuracy. This work also highlights the importance of performing multiple random trials to obtain realistic results, because a single trial could yield anywhere from 40% to 100% accuracy.

Kuehlkamp and Bowyer [4] compared the accuracy achieved using normalized and periocular iris images in gender classification. Their results showed higher accuracies for periocular iris images. Using a makeup-labeled dataset, the authors also determined that makeup can account for 2%–6% of the accuracy. Additionally, the authors performed experiments using a probabilistic occlusion mask, in which pixels are masked based on how likely they are to be considered not usable. When no mask is present, an accuracy of 64.2% is obtained. When the mask is applied to pixels that have a 70% chance of being masked, this accuracy decreases to 56.8%. The authors make note that when a single mask is used for all images, gender cues that could be present in the mask are prevented. Our paper follows the same classification method used in this work. Our main differences with Kuehlkamp and Bowyer [4] are the use of mask pairs and manual mask correction to remove gender information that could be present in the iris occlusion masks. Compared to the study by Kuehlkamp and Bowyer [4], our work further explores the impact of external information—specifically occlusion masks—through the use of mask pairing and mask correction. This aims to reduce as much external information as possible while preserving iris information.

3. Datasets and Methodology

3.1. Datasets. Two datasets were used in this work: *Gender From Iris* (GFI) [30], and *ND-CrossSensor-Iris-2013* (CSI) [47]. Both datasets were normalized using OSIRIS [44], which generated images of size 480×80 pixels and their corresponding occlusion masks. Afterward, the normalized iris images and their masks were downscaled to 240×20 and 240×40 pixels using bilinear interpolation. These resolutions are regularly used in iris processing [1, 9, 14, 30, 31]. Occlusion masks were downscaled using *max-pooling* to ensure that undesired elements remained masked in the smaller resolutions.

The first dataset we used is the GFI dataset [30]. This dataset is comprised of 3,000 iris images of size 480×640 pixels, obtained using the LG-4000 NIR sensor. These images were obtained from 1,500 subjects, 750 males and 750 females, with one image per eye. The occlusion masks generated by OSIRIS from this dataset were manually corrected using the procedure described in Section 3.3, while preserving the original masks for comparison. Afterward, the normalized iris images and their masks (original and corrected) were downscaled to 240×20 and 240×40 pixels as mentioned previously.

The second dataset we used is the CSI dataset [47]. This dataset is originally comprised of 29,986 images taken using the LG-4000 NIR sensor, and 116,564 images taken using the LG-2200 NIR sensor. These images were taken from 676 subjects, in 27 sessions spanning 3 years. From this dataset, only the images from the LG-4000 sensor were used, to reduce sources of variability. Additionally, we limited the number of images per subject to 20 (10 per eye). These images were used to verify that our method worked on a different dataset. The dimensions of the images in the second database are the same as in the first database; this is 480×640 pixels before normalization, and 240×20 or 240×40 after normalization.

3.2. Mask Pairs. As previously mentioned, it is possible that some models could be trained to classify gender using part of the mask information. It would be beneficial to eliminate this source of information when determining gender exclusively from the iris texture.

To address this, the occlusion mask distributions of both genders are equalized by generating mask pairs. The steps for generating mask pairs and equalized occlusion mask distributions are the following:

(1) Occlusion masks are grouped in male–female pairs based on a similitude criterion.
(2) The OR operation is performed on each pair, which generates a "paired mask".
(3) The newly generated paired masks are applied to the original iris images of each pair. In this way, the mask of both irises in each pair will be identical, while covering the undesired elements of each iris.

By applying the same mask to both images in each pair, the resulting mask distribution is identical for both genders. Figure 3 shows an example of this operation. The masks used in this example are not similar, so as to better illustrate the process. From now on, occlusion masks obtained using this method will be called paired masks, whereas the original masks will be called regular masks.

The first step is to group the masks in male–female pairs. In this step, every male iris is paired with a different female iris, ensuring no iris is in more than one pair. When a new paired mask is generated, some iris information can be lost in the process, as illustrated in Figure 3(b). The pixels shown in magenta are those that were already masked in both original masks, and thus will not cause loss of information. However, red pixels will add new mask pixels to the iris on the top, and blue pixels will add new mask pixels to the iris on the bottom. Pairs should be generated in a way that adds the least number of new mask pixels. This reduces the amount of iris information lost when generating and applying paired masks. To achieve this, a metric that reflects this factor will be defined.

The paired mask generated from a female mask m_f and a male mask m_m is denoted by $m_f \cup m_m$. If $|m|$ is the number of masked pixels in mask m, then the number of mask pixels that are added by pairing m_f and m_m, compared to m_f, is given by $|m_f \cup m_m| - |m_f|$. The maximum mask growth compared to both original masks will be denoted by M_G, and is calculated as follows:
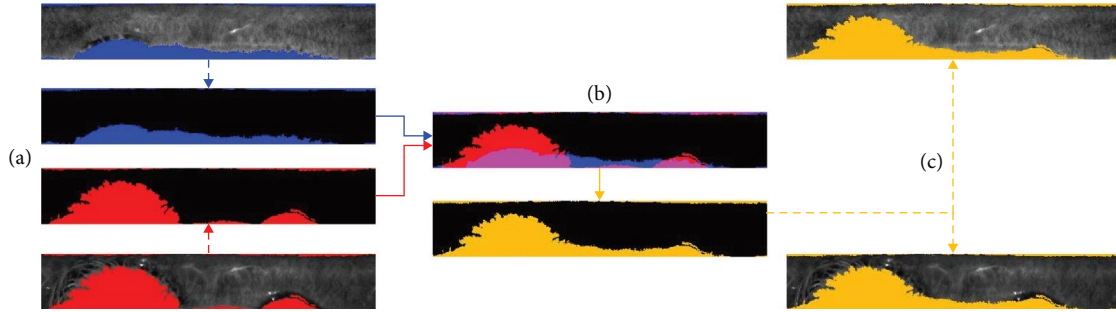
FIGURE 3: Mask pairing example. Masks with different shapes were used to better illustrate the process. (a) Masks from different-gendered iris are paired. The masks are shown in red and blue. These masks should be similar in order to reduce the information that is lost. (b) The OR operation is performed on both masks generating a paired mask, which is shown in yellow. (c) The new mask is applied to the iris.

$$M_G\big(m_f, m_m\big) = \max\left(\frac{\big|m_f \cup m_m\big| - \big|m_f\big|}{N}, \frac{\big|m_f \cup m_m\big| - \big|m_m\big|}{N}\right),$$

$$\tag{1}$$

where N is the number of pixels in each iris image. Under this definition, an (m_f, m_m) pair will be better when its $M_G(m_f, m_m)$ is lower. This value can be expressed as a percentage of the area of the iris image.

The pairs are generated by exploring every possible pair and choosing the combination that minimizes the sum of growths. To accomplish this, a matrix M_{M_G} with the M_G of every male–female pair is generated. This matrix is defined as follows:

$$M_{M_G}(i, j) = \big\{M_G\big(m_f, m_m\big)\big|m_f \in M_F, m_m \in M_M\big\}, \tag{2}$$

where M_F and M_M are the sets of all female and male iris in the dataset, respectively. With this matrix, generating the best pairs is a type of combinatorial optimization problem called assignment problem [48]. This can be efficiently solved using algorithms such as the Jonker–Volgenant algorithm [49, 50]. This algorithm finds the pairs that minimize the total M_G sum.

It is important to note that mask pairs are only required during model training; by preventing the classifier from learning gender cues present in masks during training, the classifier will not be able to use them during testing. Furthermore, in a real-life scenario, it would not be possible to generate mask pairs for previously unknown irises, since they require *a priori* knowledge of their gender.

3.2.1. Penalizing and Removing Pairs with High Growth. Despite minimizing the total M_G sum, some pairs with high M_G are generated. In these pairs, at least one of the two irises would lose considerable iris information. To study the impact of pairs with high M_G in classification, two measures were taken. The first measure consists of penalizing values in the M_{M_G} matrix above a certain threshold by multiplying them by 10^5. In this way, the algorithm will generate pairs with high M_G only if absolutely necessary. The second measure consists of removing pairs with M_G over a threshold from the dataset, which excludes them from the process.
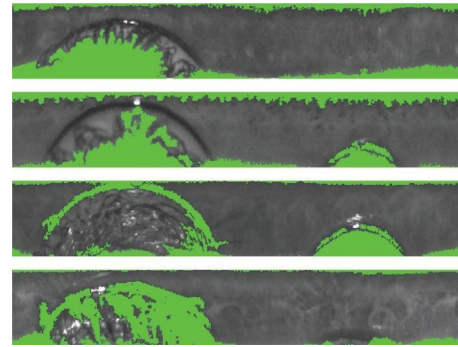

FIGURE 4: Normalized iris images. Their occlusion masks, shown in green, do not correctly cover the eyelids.

For both measures, different penalization and removal thresholds were explored. The penalization threshold was set to M_G values ranging from 3% to 15%. The removal threshold was set to M_G values ranging from 7% to 10%. This helps us to understand the impact of pairs with high M_G in classification.

3.3. Manual Mask Correction. We used OSIRIS [44] for normalizing periocular iris images. Occlusion masks generated using OSIRIS or other software often miss some non-iris areas [3]. For instance, Figure 4 shows four normalized iris images using OSIRIS, with incorrectly generated masks. Furthermore, the two images at the bottom show presence of make-up on the eyelids.

To study the effects of poorly generated masks, the masks of the GFI dataset were manually corrected using an open-source tool developed for this purpose. This tool allows a user to manually indicate the areas that should be masked while using the periocular image as a reference to discern these areas and is available for download (upon paper acceptance). Through this procedure, all non-iris elements were covered, and no pixels from the original masks were removed. Figure 5 shows two irises before and after mask correction.

3.4. Gender Classification Method

3.4.1. Features and Classifiers. Our gender classification protocols were based on those used in [4]. This involved using a pre-trained VGG-16 convolutional neural network [51] for
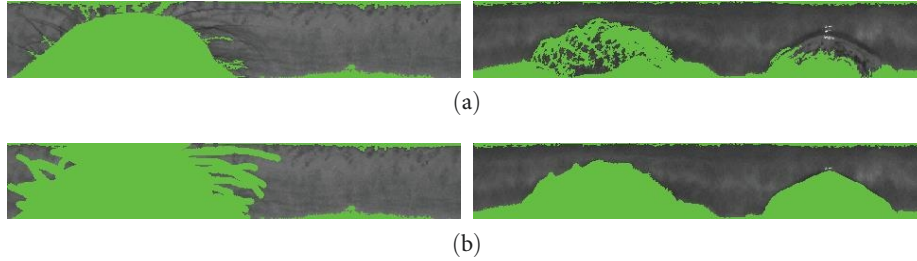
(a)

(b)

FIGURE 5: Two normalized iris images, before (a) and after (b) mask correction.

feature extraction and classification. Five different classification tests were performed. Test 1 (called "VGG-full") uses the VGG-16 network for both feature extraction and classification. In this case, we defined our own architecture for the fully connected layers, which includes three dense layers with decreasing number of neurons, and dropout layers. During training, the convolutional layers are frozen, so only the fully connected layers are adjusted.

Tests 2 through 5 use a "VGG + (classifier)" naming scheme. We used the following classifiers: linear support vector machine (LSVM) [52], support vector machine with radial basis function kernel (SVM-RBF) [52], K-nearest neighbours (KNN) [53], and random forest (RF) [54]. As input to these classifiers, 4,096 features are obtained at the output of the first fully connected layer of the VGG-16.

Tests 1 (VGG-full) and 2 (VGG + LSVM) were used for all preliminary experiments. These tests were based on those used in [4]. Tests 3 through 5 were added for comparison.

3.4.2. Preprocessing. Before entering the classifier, iris images are preprocessed as follows. The dataset is partitioned while separating subject IDs into 80% and 20% train test. The partitions are balanced to ensure each has the same number of male and female images. Masks, either regular or paired, are applied in the following manner. First, iris pixel values are restricted to the 0–254 range. Next, the median of every image is determined, and this value is assigned to the masked pixels; this prevents masked pixels from affecting the minimum or maximum values. Afterward, the image pixel values are rescaled to the 1–255 range. Finally, masked pixels are assigned the value of 0. This ensures nonmasked pixels use the whole intensity range, while preserving the 0 for masked pixels.

In the VGG + (classifier) tests, images must be resized to accommodate the VGG-16 input size. For this, all images were resized to 224×224 pixels using linear interpolation. In the VGG-full test, we were able to minimize the required resize, allowing us to keep the width intact (240 pixels), while setting the height to either 32 pixels (for the 240×20 pixels images) or the original height of 40 pixels (for the 240×40 pixels images). Since the pretrained VGG-16 network requires RGB images, our grayscale NIR images are repeated in each of the three channels as part of the preprocessing [55]. Each test was repeated using 30 different train-test partitions to ensure results are statistically significant, as done in [4].

When applying masks, we have the option to use either regular masks or paired masks. Additionally, for the GFI dataset, we have the option to use either the original (OSIRIS) masks, or to use the manually corrected masks. Together with the mask pairs, this yields a total of four possible mask combinations. In the case of paired masks, these are only applied to the train partition, as mentioned in Section 3.2.

4. Results and Discussion

In this section, the results of the proposed mask pairing method are shown and discussed.

Section 4.1 illustrates how the mask distributions per gender changed after manually correcting and pairing. Section 4.2 describes the effects of changing penalization thresholds on pair distribution and gender classification. Section 4.3 shows the effects of changing removal thresholds on gender classification. Finally, Section 4.4 summarizes the classification results, and the effects of mask pairs on gender classification.

4.1. Effects of Pairing on Mask Distribution. The original distribution of masks per gender, for both manually corrected and original masks, is shown in Figure 6. The areas without overlap indicate differences in mask distribution among both genders.

After generating pairs using all the images from the training partition, mask distribution in this partition becomes identical for both genders. This is reflected in Figure 7, in which the histograms completely overlap. Compared to Figure 6, there are more images with a higher percentage of mask, because pairing adds mask to the iris.

4.2. Effects of Pair Penalization. As described in Section 3.2.1, different thresholds for penalizing and removing pairs were compared. To select the right penalization threshold, two elements were analyzed: average mask growth of generated pairs and number of pairs with high (>10%) growth. A threshold will be better the lower both these numbers are. Figure 8 shows the distribution of pairs at different penalization thresholds. This figure shows that, as the threshold decreases and becomes stricter, the average growth increases until it reaches its maximum between thresholds 5% and 7.5%. Beyond this point, because most potential pairs have been penalized, the behavior reverses as if less pairs had been penalized. The number of pairs with high growth reaches a minimum at a threshold of 10%. This is in part caused by defining "pair with high growth" at the same percentage, as this is the least strict threshold that will prevent pairs of high growth from being generated.
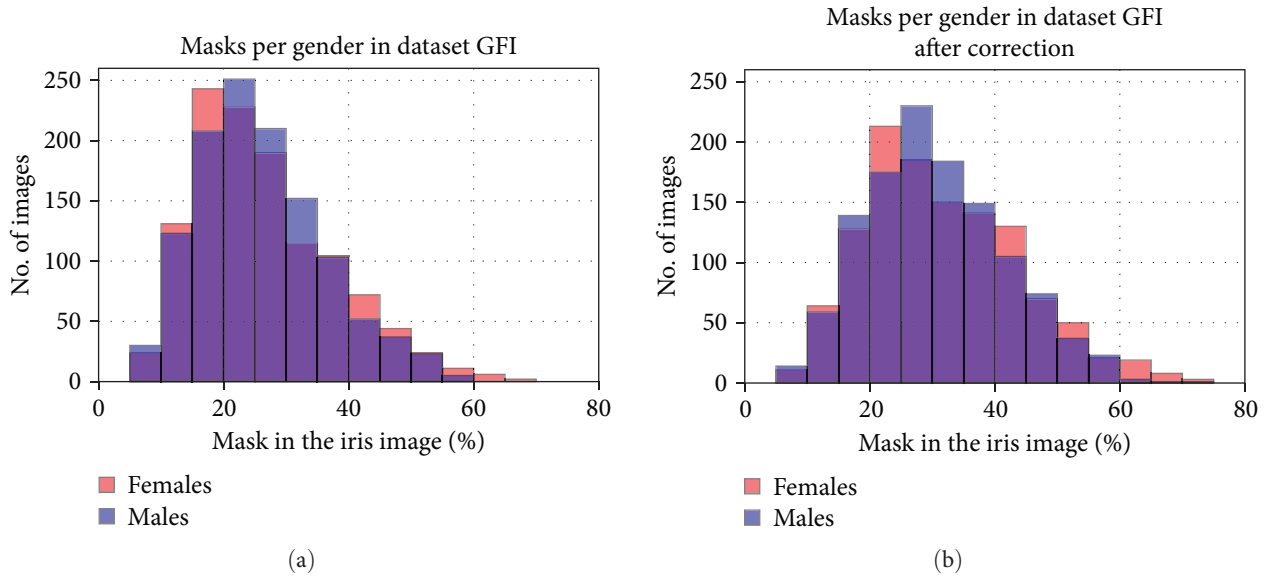
FIGURE 6: Mask distribution for each gender, (a) before generating pairs, and using original masks or (b) corrected masks. Columns are shown in purple when both genders are overlapped.
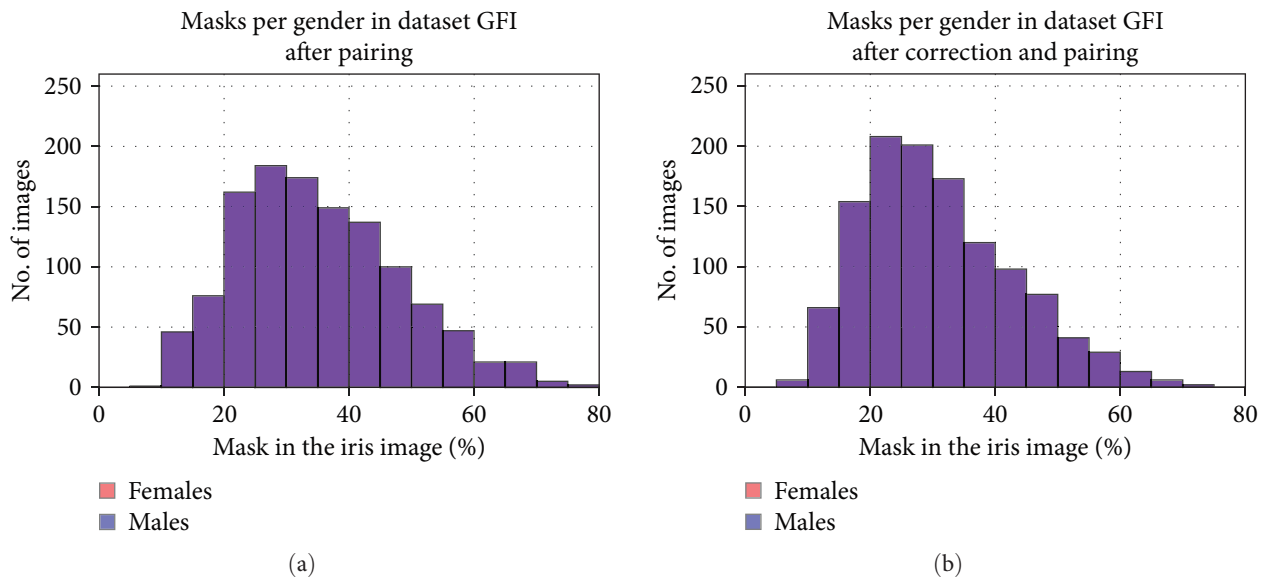

FIGURE 7: Mask distribution for each gender, (a) after generating pairs, and using original masks or (b) corrected masks. Columns are shown in purple when both genders are overlapped.

With regards to the impact in classification results, every threshold was used in five trials for each resolution. An ANOVA test was performed to assess if changing the penalization threshold has any significant effect on the classification accuracy. The results are shown in Figure 9. When using the original masks, the mean accuracy ranged from 59.14% to 60.59%. When using the manually corrected masks, the mean accuracy ranged from 56.37% to 57.20%. In both cases the effect of changing the penalization threshold was not significant (p-value close to 1). From these results and the pair distribution results, a penalization threshold of 10% was chosen.

4.3. Effects of Pair Removal. The effects of pair removal in classification were studied. An ANOVA test was performed to evaluate whether the changes in removal threshold have any significant effect on the classification accuracy. Results are shown in Table 2. The p-values indicate that the effect was not significant. In conclusion, removing pairs with high growth has no impact on classification.

4.4. Classification Results. Our classification results were summarized in Table 3 for the GFI dataset, and in Table 4 for the CSI dataset. GFI results show an average decrease in
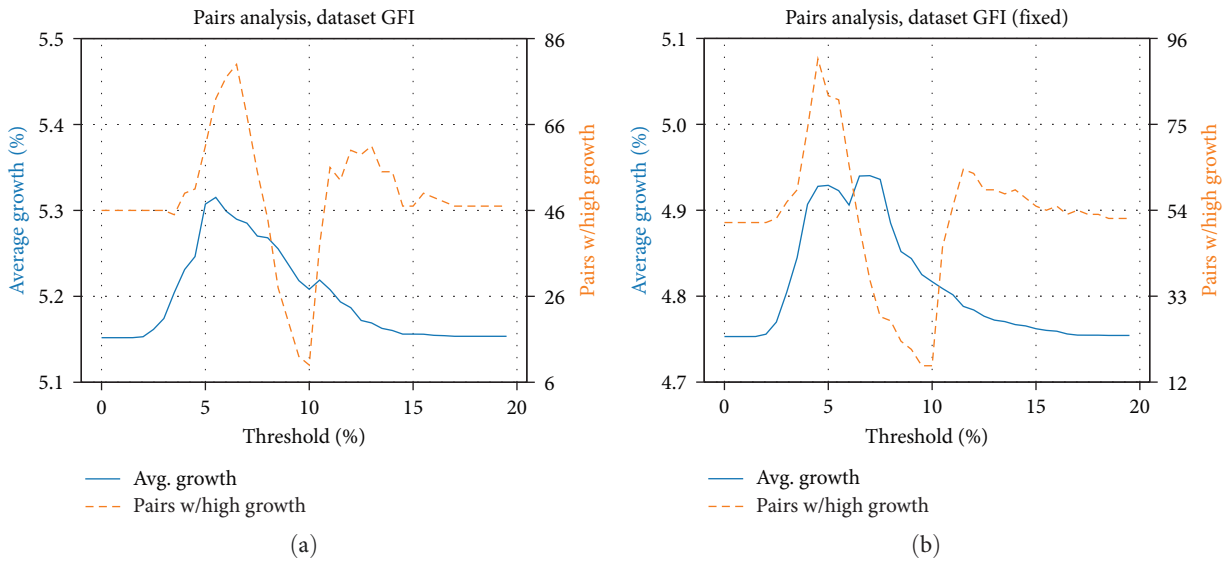
FIGURE 8: Average mask growth and number of pairs with high growth, for different penalization thresholds, and (a) using original masks or (b) corrected masks.
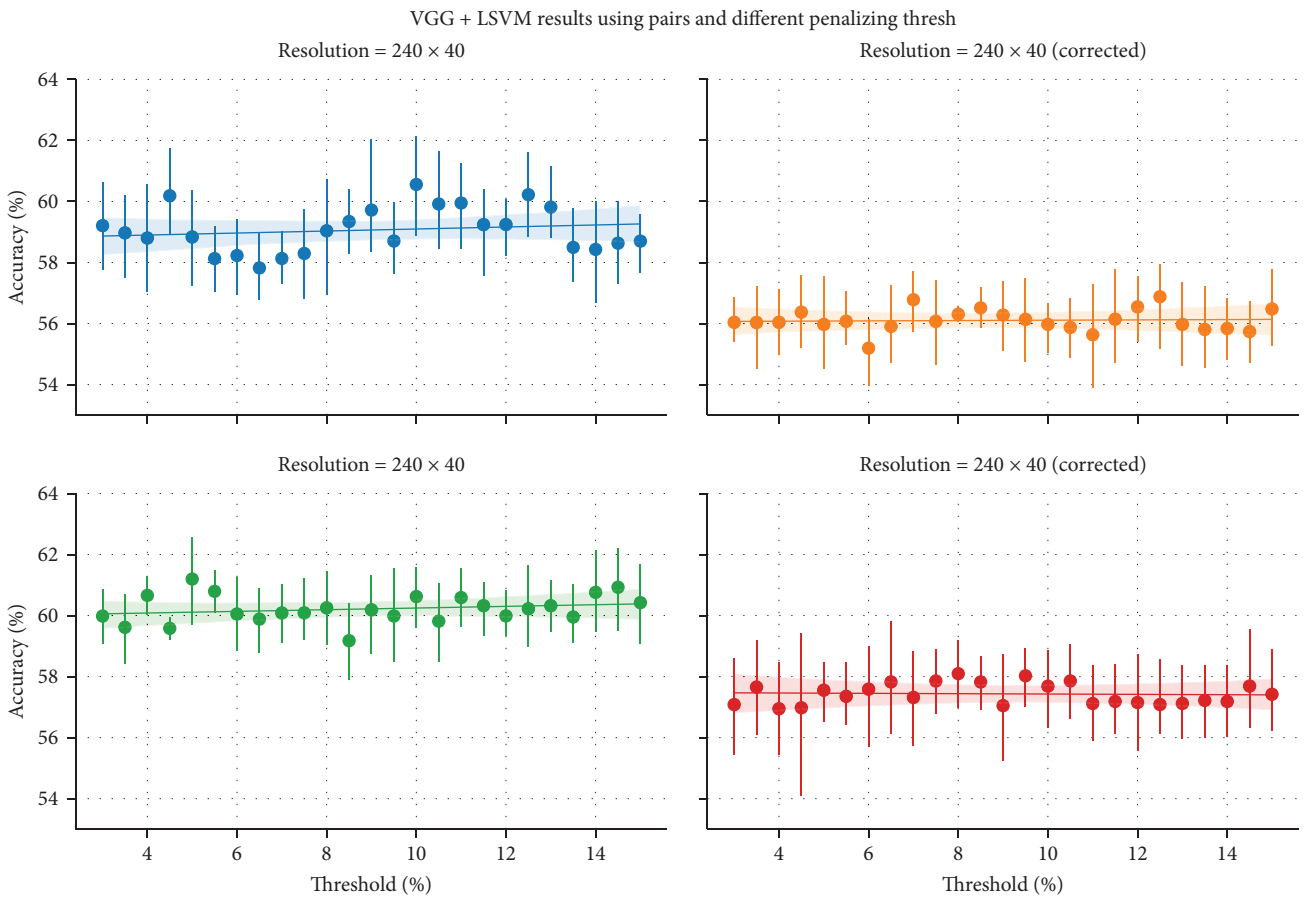

FIGURE 9: Classification results at different resolutions, using different penalization thresholds. Pairs over the threshold are penalized during training. Resolutions labeled as "(corrected)" used corrected masks instead of original masks. A linear regression is shown alongside the results.

TABLE 2: Classification results after removing pairs over a certain growth threshold.

Test type	Resolution (px)	Threshold (%)	Accuracy (%)	p-Value
VGG-full	240 × 20	10	**58.38 ± 2.59**	0.78
		9	57.64 ± 2.93	
		8	57.78 ± 3.32	
		7	57.74 ± 3.37	
	240 × 40	10	59.61 ± 2.91	0.99
		9	59.54 ± 2.86	
		8	**59.72 ± 3.07**	
		7	59.54 ± 2.83	
VGG + LSVM	240 × 20	10	58.66 ± 2.16	0.97
		9	58.83 ± 2.19	
		8	58.72 ± 2.35	
		7	**58.90 ± 2.09**	
	240 × 40	10	60.48 ± 2.03	0.98
		9	**60.64 ± 1.98**	
		8	60.56 ± 2.06	
		7	60.41 ± 1.97	

Highest and lowest accuracies are shown in bold letters.

TABLE 3: Results using the GFI dataset, showing the effect of using paired and corrected masks.

Test type	Paired masks	Resolution	Corrected masks	Accuracy (%)
VGG-full	No	240 × 20	No	62.13 ± 2.37
			Yes	61.69 ± 2.72
		240 × 40	No	**63.34 ± 2.64**
			Yes	61.89 ± 2.95
	Yes	240 × 20	No	58.36 ± 2.84
			Yes	56.39 ± 3.36
		240 × 40	No	59.11 ± 3.10
			Yes	**55.99 ± 2.82**
VGG + LSVM	No	240 × 20	No	63.97 ± 1.64
			Yes	62.79 ± 1.67
		240 × 40	No	**65.43 ± 1.93**
			Yes	62.01 ± 2.12
	Yes	240 × 20	No	58.72 ± 2.02
			Yes	**56.42 ± 1.77**
		240 × 40	No	60.62 ± 2.13
			Yes	56.96 ± 1.69
VGG + SVM-RBF	No	240 × 20	No	63.94 ± 2.27
			Yes	63.24 ± 2.18
		240 × 40	No	**64.09 ± 2.50**
			Yes	62.74 ± 1.87
	Yes	240 × 20	No	58.02 ± 2.21
			Yes	56.14 ± 1.79
		240 × 40	No	60.46 ± 2.01
			Yes	**55.83 ± 1.85**
VGG + KNN	No	240 × 20	No	57.17 ± 1.55
			Yes	56.85 ± 1.55
		240 × 40	No	**57.63 ± 1.85**
			Yes	55.79 ± 1.62
	Yes	240 × 20	No	52.81 ± 1.85
			Yes	51.98 ± 1.62
		240 × 40	No	54.15 ± 1.71
			Yes	**51.93 ± 1.89**

TABLE 3: Continued.

Test type	Paired masks	Resolution	Corrected masks	Accuracy (%)
VGG + RF	No	240 × 20	No	60.64 ± 2.31
			Yes	**61.29 ± 1.93**
		240 × 40	No	61.28 ± 2.10
			Yes	60.46 ± 2.31
	Yes	240 × 20	No	55.90 ± 2.08
			Yes	**54.37 ± 1.77**
		240 × 40	No	56.55 ± 1.73
			Yes	54.44 ± 1.94

Highest and lowest accuracies for each test are shown in bold letters.

TABLE 4: Results using the CSI dataset, showing the effect of using paired masks.

Test type	Paired masks	Resolution	Accuracy (%)
VGG-full	No	240 × 20	59.90 ± 2.68
		240 × 40	**60.13 ± 3.05**
	Yes	240 × 20	**56.54 ± 2.28**
		240 × 40	57.18 ± 2.95
VGG + LSVM	No	240 × 20	**61.44 ± 2.71**
		240 × 40	61.10 ± 2.54
	Yes	240 × 20	57.44 ± 2.82
		240 × 40	**57.25 ± 2.60**
VGG + SVM-RBF	No	240 × 20	**62.34 ± 3.12**
		240 × 40	62.36 ± 3.07
	Yes	240 × 20	58.37 ± 3.14
		240 × 40	**58.96 ± 2.92**
VGG + KNN	No	240 × 20	**56.70 ± 2.08**
		240 × 40	56.46 ± 1.73
	Yes	240 × 20	52.66 ± 1.65
		240 × 40	**52.50 ± 1.43**
VGG + RF	No	240 × 20	**59.42 ± 2.71**
		240 × 40	59.90 ± 2.55
	Yes	240 × 20	56.15 ± 2.55
		240 × 40	56.31 ± 2.44

Highest and lowest results for each test are shown in bold letters.

accuracy of 5.16% when using mask pairs, whereas using corrected masks decreases accuracy by 1.76% on average. Using both paired and corrected masks decreases accuracy by 6.92% on average. CSI results also show an average decrease in accuracy of 3.64%. We performed ANOVA tests to evaluate the impact of using paired masks and corrected masks, and both were shown to be significant ($p < 0.01$). This demonstrates that occlusion masks introduce significant gender information, and that the actual gender information in the iris texture is less than the previously reported.

These results are consistent with those obtained in [4], on the GFI dataset. The accuracy in their equivalent VGG + LSVM test was 60.0%, and in their equivalent VGG-full test was 60.1%. Our nonpaired results were slightly better, whereas our paired results were slightly worse. This means that despite having higher baseline results, using paired masks reduced the gender classification accuracy significantly.

However, our results are not consistent with the previous papers that reported accuracies of over 70% when using normalized iris [30–32]. Nevertheless, none of these papers account for the bias that the masks may introduce.

The information provided by the masks can potentially contribute to gender classification as it may exhibit correlations with gender through factors such as makeup, the segmentation algorithm employed, and the specific conditions encountered during image acquisition. Therefore, if the objective is to perform gender classification using the iris texture exclusively, it is essential to remove this information.

When using mask pairs, gender information can be lost in two ways: (1) through the gender cues present in the occlusion masks, and (2) through the iris pixels covered by the paired masks. However, we have yet to quantify the amount of gender information lost through these covered iris pixels. Interestingly, we found that removing the worst

mask pairs and using stricter penalization thresholds did not lead to a decrease in accuracy. This suggests that the information lost through the masks is greater than the information lost through the masked iris pixels. Nevertheless, we still need to demonstrate that the iris information covered by the paired masks does not contain important gender cues.

5. Conclusions

In this work, we proposed a novel method for studying gender information exclusively in the iris by removing gender information that could be present in the iris occlusion masks. This is achieved by pairing masks according to gender. Using this method, gender classification accuracy decreased by 4.65% on average on the GFI and CSI datasets. Furthermore, when using both corrected and paired masks, accuracy decreased by 6.92% on average. This demonstrates that occlusion masks introduce significant gender information, and that iris classification results from the other works do not reflect the information exclusively in the iris but also in the occlusion masks and non-iris elements. Using iris information exclusively may yield significantly lower accuracy than the previously reported.

Pairing the masks eliminates possible correlation between gender and mask shape. Our results showed that the loss in accuracy after applying our method is significative. Therefore, works aiming to perform gender classification using the iris texture from normalized iris images should eliminate this correlation.

Future work could focus on quantifying the information loss while generating mask pairs.

If any such information exists and is significant, a methodology that recovers this information without reintroducing skews would be of interest. Additionally, using a larger dataset where presence of makeup is annotated (such as in [4]) could provide better insight on the impact of masks and makeup in the gender classification.

Newer normalization and segmentation techniques utilize deep learning for the purpose of processing iris images [56, 57]. These techniques could be utilized instead of OSIRIS, as they perform more accurate segmentation, which would make manual mask correction unnecessary. This improved segmentation could also reduce the amount of iris information that is lost when generating pairs.

Data Availability

GFI and CSI datasets were provided by the University of Notre Dame and are available for request at the following URL: https://cvrl.nd.edu/projects/data/. The toolbox for manually correcting occlusion masks is freely available at https://github.com/Nosferath/fixMasks. Other algorithms, code, and results may be requested to Claudio Yáñez, claudio.yanez@ug.uchile.cl.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] J. Tapia, C. A. Perez, and K. Bowyer, "Gender classification from iris images using fusion of uniform local binary patterns," in *Computer Vision—ECCV 2014 Workshops. ECCV 2014*, L. Agapito, M. Bronstein, and C. Rother, Eds., vol. 8926 of *Lecture Notes in Computer Science*, pp. 751–763, Springer, Cham, 2015.

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

[3] Y.-H. Li and M. Savvides, "An automatic iris occlusion estimation method based on high-dimensional density estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 4, pp. 784–796, 2013.

[4] A. Kuehlkamp and K. Bowyer, "Predicting gender from iris texture may be harder than it seems," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 904–912, IEEE, Waikoloa, HI, USA, 2019.

[5] J. R. Malgheet, N. B. Manshor, and L. S. Affendey, "Iris recognition development techniques: a comprehensive review," *Complexity*, vol. 2021, Article ID 6641247, 32 pages, 2021.

[6] A. Makrushin, A. Uhl, and J. Dittmann, "A survey on synthetic biometrics: fingerprint, face, iris and vascular patterns," *IEEE Access*, vol. 11, pp. 33887–33899, 2023.

[7] V. S. Kumar, N. B. Reddy, and U. Kiirran, "Predictive analytics on gender classification using machine learning," in *2023 10th International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 1634–1639, IEEE Xplore, New Delhi, India, March 2023.

[8] V. K. Suravarapu and H. Y. Patil, "Person identification and gender classification based on vision transformers for periocular images," *Applied Sciences*, vol. 13, no. 5, Article ID 3116, 2023.

[9] D. Bobeldyk and A. Ross, "Iris or periocular? Exploring sex prediction from near infrared ocular images," in *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*, IEEE, Denver, CO, USA, 2016.

[10] J. E. Tapia and I. Viedma, "Gender classification from multispectral periocular images," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pp. 805–812, IEEE, Denver, CO, USA, 2017.

[11] I. Viedma and J. E. Tapia, "Deep gender classification and visualization of near-infra-red periocular-iris images," in *2018 IEEE International Conference on Image Processing, Applications and Systems (IPAS)*, pp. 73–78, IEEE, Sophia Antipolis, France, December 2018.

[12] D. Bobeldyk and A. Ross, "Predicting gender and race from near infrared iris and periocular images," 2018, CoRR, http://arxiv.org/abs/1805.01912.

[13] M. Larsson and N. L. Pedersen, "Genetic correlations among texture characteristics in the human iris," *Molecular Vision*, vol. 10, pp. 821–831, 2004.

[14] A. Kuehlkamp, B. Becker, and K. Bowyer, "Gender-from-iris or gender-from-mascara?" in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1151–1159, IEEE, Santa Rosa, CA, 2017.

[15] M. Singh, S. Nagpal, M. Vatsa, R. Singh, A. Noore, and A. Majumdar, "Gender and ethnicity classification of iris images using deep class-encoder," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pp. 666–673, IEEE, Denver, CO, USA, 2017.

[16] A. Rattani, N. Reddy, and R. Derakhshani, "Gender prediction from mobile ocular images: a feasibility study," in *2017 IEEE International Symposium on Technologies for Homeland Security (HST)*, pp. 1–6, IEEE, Waltham, MA, USA, April 2017.

[17] A. Rattani, N. Reddy, and R. Derakhshani, "Convolutional neural networks for gender prediction from smartphone-based ocular images," *IET Biometrics*, vol. 7, no. 5, pp. 423–430, 2018.

[18] J. Tapia and C. Aravena, "Gender classification from periocular NIR images using fusion of CNNs models," in *2018 IEEE 4th International Conference on Identity, Security, and Behavior Analysis (ISBA)*, pp. 1–6, IEEE, Singapore, January 2018.

[19] D. Bobeldyk and A. Ross, "Predicting soft biometric attributes from 30 pixels: a case study in NIR ocular images," in *2019 IEEE Winter Applications of Computer Vision Workshops (WACVW)*, pp. 116–124, IEEE, Waikoloa, HI, USA, January 2019.

[20] A. Manyala, H. Cholakkal, V. Anand, V. Kanhangad, and D. Rajan, "CNN-based gender classification in near-infrared periocular images," *Pattern Analysis and Applications*, vol. 22, no. 4, pp. 1493–1504, 2019.

[21] I. Viedma, J. Tapia, A. Iturriaga, and C. Busch, "Relevant features for gender classification in NIR periocular images," *IET Biometrics*, vol. 8, no. 5, pp. 340–350, 2019.

[22] J. Tapia, C. Rathgeb, and C. Busch, "Sex-prediction from periocular images across multiple sensors and spectra," in *2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pp. 529–535, IEEE, Spain, 2018.

[23] J. E. Tapia and C. Arellano, "Soft-biometrics encoding conditional GAN for synthesis of NIR periocular images," *Future Generation Computer Systems*, vol. 97, pp. 503–511, 2019.

[24] J. Tapia, C. Arellano, and I. Viedma, "Sex-classification from cellphones periocular iris images," in *Selfie Biometrics: Advances and Challenges*, pp. 227–242, Springer International Publishing, Cham, 2019.

[25] M. Eskandari and O. Sharifi, "Effect of face and ocular multimodal biometric systems on gender classification," *IET Biometrics*, vol. 8, no. 4, pp. 243–248, 2019.

[26] F. Alonso-Fernandez, K. H. Diaz, S. Ramis, F. J. Perales, and J. Bigun, "Soft-biometrics estimation in the era of facial masks," in *2020 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pp. 1–6, IEEE, Darmstadt, Germany, 2020.

[27] K. B. Raja, R. Raghavendra, and C. Busch, "Fused spectral features in kernel weighted collaborative representation for gender classification using ocular images," in *Proceedings of 3rd International Conference on Computer Vision and Image Processing*, B. B. Chaudhuri, M. Nakagawa, P. Khanna, and S. Kumar, Eds., pp. 131–143, vol. 1022 of *Advances in Intelligent Systems and Computing*, Springer, Singapore, 2020.

[28] V. Thomas, N. V. Chawla, K. W. Bowyer, and P. J. Flynn, "Learning to predict gender from iris images," in *2007 First IEEE International Conference on Biometrics: Theory, Applications, and Systems*, pp. 1–5, IEEE, Crystal City, VA, USA, 2007.

[29] A. Bansal, R. Agarwal, and R. K. Sharma, "SVM based gender classification using iris images," in *2012 Fourth International Conference on Computational Intelligence and Communication Networks*, pp. 425–429, IEEE, Mathura, India, 2012.

[30] J. E. Tapia, C. A. Perez, and K. W. Bowyer, "Gender classification from the same iris code used for recognition," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 8, pp. 1760–1770, 2016.

[31] J. Tapia and C. Aravena, "Gender classification from nir iris images using deep learning." in *Deep Learning for Biometrics*, Advances in Computer Vision and Pattern Recognition, pp. 219–239, Springer, Cham, 2017.

[32] J. E. Tapia and C. A. Perez, "Gender classification from NIR images by using quadrature encoding filters of the most relevant features," *IEEE Access*, vol. 7, pp. 29114–29127, 2019.

[33] G. Sutra, S. Garcia-Salicetti, and B. Dorizzi, "The Viterbi algorithm at different resolutions for enhanced iris segmentation," in *Proceedings—2012 5th IAPR International Conference on Biometrics (2012)*, pp. 310–316, IEEE, India, 2012.

[34] J. W. Thompson and P. J. Flynn, "A segmentation perturbation method for improved iris recognition," in *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pp. 1–8, 2010.

[35] M. Da Costa-Abreu, M. Fairhurst, and M. Erbilek, "Exploring gender prediction from iris biometrics," in *2015 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pp. 1–11, IEEE, Darmstadt, Germany, September 2015.

[36] J. Tapia and C. Arellano, "Gender classification from iris texture images using a new set of binary statistical image features," in *2019 International Conference on Biometrics (ICB)*, pp. 1–7, IEEE, Greece, 2019.

[37] G. S. Sable and M. R. Rajput, "Radon and multiwavelet-based compact feature vector generation for gender identification from iris," in *Advances in Intelligent Systems and Computing*, vol. 1025, pp. 29–43, Springer, 2020.

[38] F. Alonso-Fernandez, K. Hernandez-Diaz, S. Ramis, F. J. Perales, and J. Bigun, "Facial masks and soft-biometrics: leveraging face recognition CNNs for age and gender prediction on mobile ocular images," *IET Biometrics*, vol. 10, no. 5, pp. 562–580, 2021.

[39] N. Vetrekar, R. Ramachandra, K. Raja, and R. Gad, "Gender classification under eyeglass occluded ocular region: an extensive study using multi-spectral imaging," in *AI and Deep Learning in Biometric Security*, pp. 175–203, CRC Press, 2021.

[40] B. Patil and M. Hangarge, "Gender prediction using iris features," in *Emerging Research in Computing, Information, Communication and Applications*, vol. 928, pp. 587–595, Springer, Singapore, 2023.

[41] R. Alrabiah, M. Hussain, and H. AboAlSamh, "Unconstrained gender recognition from periocular region using multiscale deep features," *Intelligent Automation & Soft Computing*, vol. 35, no. 3, pp. 2941–2962, 2023.

[42] L. Masek, *Recognition of human iris patterns for biometric identification*, [Unpublished thesis], The University of Western Australia, 2003.

[43] X. Liu, K. W. Bowyer, and P. J. Flynn, "Experiments with an improved iris segmentation algorithm," in *Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05)*, pp. 118–123, IEEE, October 2005.

[44] N. Othman, S. Guillaume, B. Dorizzi, and S. Garcia-Salicetti, *A Biometric Reference System for Iris OSIRIS Version 4.1*, BioSecure, Telecom Sud Paris, 2013.

[45] N. Othman, B. Dorizzi, and S. Garcia-Salicetti, "OSIRIS: an open source iris recognition software," *Pattern Recognition Letters*, vol. 82, pp. 124–131, 2016.

[46] K. P. Hollingsworth, K. W. Bowyer, and P. J. Flynn, "The best bits in an iris code," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 6, pp. 964–973, 2009.

[47] N. Popescu-Bodorin, L. S. Grigore, V. E. Balas et al., "Cross-sensor iris recognition: LG4000-to-LG2200 comparison," arXiv preprint. arXiv: 180101695. 2013

[48] R. Burkard, M. Dell'Amico, and S. Martello, *Assignment Problems*, SIAM Publishing Library, Philadelphia, PA, 2009.

[49] R. Jonker and A. Volgenant, "A shortest augmenting path algorithm for dense and sparse linear assignment problems," *Computing*, vol. 38, no. 4, pp. 325–340, 1987.

[50] D. F. Crouse, "On implementing 2D rectangular assignment algorithms," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 4, pp. 1679–1696, 2016.

[51] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations (ICLR 2015)*, pp. 1–14, Computational and Biological Learning Society, 2015.

[52] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[53] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[54] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[55] J. E. Zambrano, D. P. Benalcazar, C. A. Perez, and K. W. Bowyer, "Iris recognition using low-level CNN layers without training and single matching," *IEEE Access*, vol. 10, pp. 41276–41286, 2022.

[56] A. Valenzuela, C. Arellano, and J. E. Tapia, "Towards an efficient segmentation algorithm for near-infrared eyes images," *IEEE Access*, vol. 8, pp. 171598–171607, 2020.

[57] E. Jalilian, M. Karakaya, and A. Uhl, "End-to-end off-angle iris recognition using CNN based iris segmentation," in *2020 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pp. 1–7, IEEE Xplore, Darmstadt, Germany, September 2020.