Gender Classification from the Same Iris Code Used for Recognition

Juan E. Tapia, Student Member, IEEE Claudio A. Perez, Senior Member, IEEE Kevin W. Bowyer Fellow Member, IEEE

Abstract—Previous researchers have explored various approaches for predicting the gender of a person based on features of the iris texture. This paper is the first to predict gender directly from the same binary iris code that could be used for recognition. We found that information for gender prediction is distributed across the iris, rather than localized in particular concentric bands. We also found that using selected features representing a subset of the iris region achieves better accuracy than using features representing the whole iris region. We used measures of mutual information to guide the selection of bits from the iris code to use as features in gender prediction. Using this approach, with a person-disjoint training and testing evaluation, we were able to achieve 89% correct gender prediction using the fusion of the best features of iris code from the left and the right eyes.

Index Terms—Gender Classification, Iris, Feature Selection.

I. INTRODUCTION

One active area of ‘soft biometrics’ research involves classifying the gender of the person from the biometric sample. Most work done on gender classification has involved the analysis of face images [1]. Various types of feature extraction, selection and classifiers have been used in gender classification.

In terms of matching iris codes for recognition of identity, iris codes of different individuals, and even of the left and right eyes of the same individual, have been shown to be independent. At the same time, several authors have reported that, using an analysis of iris texture different from that used for recognition of identity, it is possible to classify the gender of the person with accuracy much higher than chance.

Nowadays, essentially all commercial systems to identify people from iris are based on the iris-code proposed by Daugman [2]. Therefore, the iris-code is already being computed in iris recognition systems and could be used for other purposes such as gender prediction, either to help speed the matching process, and / or to know something about people who are not recognized. Commercial iris recognition systems typically do not also acquire face images or fingerprint images, and so gender-from-iris is the only option for gender info in an iris recognition system. Our approach is the first to classify gender from the same iris-code used to identify people. If the gender is computed before a search for a match to an enrolled iris code, then the average search time can potentially be cut in half. In instances where the person is not recognized, it may be useful to know the gender and other information about people trying to gain entry.

There are a number of reasons why ‘gender-from-iris’ is an interesting and potentially useful problem. One possible use arises in searching an enrolled database for a match. If the gender of the sample can be determined, then it can be used to order the search and reduce the average search time. Another possible use arises in social settings where it may be useful to screen entry to some area based on gender, but without recording identity. Gender classification is also important for demographic information collection, marketing research, and real time electronic marketing. For example, displays at retail stores could offer products according to the person’s gender. Another possible use is that in high-security scenarios, there may be value to knowing the gender of the persons who attempt entry but are not recognized as any of the enrolled persons. And, at a basic science level, it is of value to more fully understand what information about a person can be extracted from analysis of their iris texture.

Gender classification using iris information is a rather new topic, with only a few papers published [3], [4], [5], [6]. Most gender classification methods reported in the literature use texture features from the whole iris for classification or identification [3], [4], [5]. As a result, gender-irrelevant information might be fed into the classifier which may result in poor generalization.

Thomas et al. [5] were the first to explore gender-from-iris problem. They used a set of over 50,000 left-iris images, and performed several different experiments. They segmented the iris region, created a normalized iris image, and then created a log-Gabor filtered version of the normalized image. In addition to the log-Gabor texture features, they used seven geometric features of the pupil and iris. They developed a random subspace ensemble of decision trees to predict gender based on the iris texture and the geometric features, and reported accuracy close to 80% in certain circumstances.

Lagree et al. [4] computed texture features separately for eight five-pixel horizontal bands, running from the pupil-iris boundary out to the iris-sclera boundary, and ten twenty-four-pixel vertical bands from a 40x240 image. The normalized image is not processed by the log-Gabor filters that are used to create the iris code that is used for identity recognition purposes. Also, this work does not use any geometric features such as used in [5]. Classifiers are developed to predict gender and ethnicity based on the texture features computed from the normalized iris image. Using a dataset of 600 images
and representing 60 persons, they reported achieving accuracy close to 62% for gender prediction, and close to 80% for two-class (Asian, Caucasian) ethnicity prediction.

Bansal et al. [3] used a statistical feature extraction technique based on correlation between adjacent pixels, that was combined with a 2D wavelet tree based on feature extraction techniques to extract significant features from the iris image. This approach yielded an accuracy of 83.06% for gender classification. However, the database used in this experiment was small (300 images, and without person-disjoint training) compared to other studies.

Tapia et al. [6] explored using different implementations of Local Binary Patterns (LBP) from the normalized iris image masked for iris region occlusion. They found that Uniform LBP with concatenated histograms significantly improved accuracy of gender prediction relative to using the whole iris image. Although the results in [6] using 1,500 images were intended to be computed on the basis of a person-disjoint training and testing division of the dataset, this was not actually achieved due to the dataset having multiple images for some subjects, with an average of about six images per subject. The highest accuracy achieved with any of the LBP variants was just over 91% correct gender prediction. This current paper improves on the work in [6] in several ways. One, the texture features for gender prediction are taken directly from the binary iris code that could be used for identity recognition. Two, this paper uses a new dataset that has only one image per iris, making it possible to compute performance for a truly person-disjoint train and test split of the data and the proposed feature selection method selects groups of features instead of pairs as done by traditional methods.

None of the previous gender-from-iris work has attempted to predict gender from the same binary iris code that is computed for recognizing/verifying identity. The various approaches have each computed their own different type of texture feature from the iris image or a Gabor-filtered version of this image. Also, most of the approaches have used features computed from the whole iris region, rather than selecting features representing a subset of iris area to try to maximize accuracy. Hollingsworth et al. [7] were the first to present an experiment documenting that some bits are more consistent than others. While their work was done in the context of person recognition rather than gender classification, the idea that not all bits of the iris code are equally useful may apply to both problems. The fragile-bits results suggest that using a selected subset of the iris, rather than the whole iris region, may improve accuracy.

We propose a new method to select a subset of the binary iris code to improve gender classification. (A bit position within the iris code corresponds to a particular location on the iris). We first conduct a simple experiment to determine whether information for gender prediction is localized in only some bands of the iris, See Figure 1. We find that information for gender prediction appears to be present throughout the iris. We then consider feature selection methods based on Mutual Information (MI) in order to select a subset of the available features. We explore four feature-selection methods based on MI as a measure of relevance and redundancy among features: minimal Redundancy and Maximal Relevance (mRMR) and Conditional Mutual Information Maximization (CMIM) using pairs of features, and two improved methods Weighted mRMR (W-mRMR) and Weighted CMIM (W-CMIM) using groups of features. Using the weighted CMIM (W-CMIM) approach to select the best features (bits) present in the iris code, we were able to achieve over 89% correct gender classification, computed on a person-disjoint training and testing set.

The main novelties of the work reported here are: (a) to find a new high accuracy for gender-from-iris, (b) to show that selecting a subset of possible features improves performance, and (c) to show that person-disjoint methodology is needed to make a realistic performance estimate.

The remainder of the paper is organized as follows. Section 2 reviews background concepts in iris recognition and information theory, and explains our proposed method. Section 3 details the dataset used in our experiments. Section 4 describes the experiments and results. Finally, conclusions and discussion are given in Section 5.

II. METHODS

A. Feature Extraction

The iris feature extraction process involves the following steps. First, the iris sensor acquires a near-infrared image of the eye. Near-infrared illumination allows iris texture to be imaged in both dark and light eyes. Then the iris region is segmented, and transformed into a fixed-size rectangular 'unwrapped' iris image. This facilitates creating a fixed-size iris code that is easily compared to other codes.

A texture filter is applied at a grid of locations on this unwrapped iris image, and the filter responses are quantized to yield a binary iris code [8]. Our method is intended to find gender from the iris code because the iris code is increasingly being used in large-scale identity verification systems. Iris classification has become increasingly important in large populations as in UAE, Canada, and India in recent years [9], [10], [11], and therefore the possibility to classify gender from iris is a reality. The radial resolution ($r$) and angular resolution ($\theta$) used during the normalization or 'unwrapping' step determine the size of the rectangular iris image, and can significantly influence the iris recognition rate. In this work we use a rectangular image of $20(r) \times 240(\theta)$ created using the IrisBee implementation [12]. IrisBee also creates a segmentation mask of the same size as the rectangular image. The segmentation mask indicates the portions of the normalized iris image that represent occlusion by eyelids, eyelashes or specular reflections.

The 2D normalized pattern is broken up into a number of 1D signals, and then these 1D signals are convolved with 1D Gabor filters with the following parameters: the wavelength (in pixels) of the log Gabor filter was 18 pixels and the ratio ($\sigma/f_0$) of the log Gabor filter was 0.5, where $\sigma$ denotes bandwidth and $f_0$ denotes the central frequency [12]. The output of the Gabor filters is transformed into the binary iris code by quantizing the phase information into four levels, representing the four quadrants in the complex plane [2].
A feature is statistically relevant if its removal from a feature set will reduce the prediction power \[14\]. Some researchers propose to remove redundant features from the feature list, as this may improve the prediction accuracy \[15\]. Other researchers noticed that the removal of the redundant features may cause the exclusion of potential relevant features \[14\]. Therefore, they propose to find surrogate features by measuring feature correlations, or group features with similar patterns into feature clusters. Thus, we use Mutual Information to measure the relationship between features.

MI is defined as a measure of how much information is jointly contained in two variables \[16\], \[17\], or the degree to which knowledge of one variable determines the other variable. It forms the basis of information-theoretic feature selection, as it provides a function for calculating the relevance of a variable with respect to the target class.

\[MI(x, y) = \int \int p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dxdy\]  

(1)

We use categorical \(MI\) in this paper, which can be estimated by tallying the samples of categorical variables in the data building adaptive histograms to compute the joint probability distribution \(p(x,y)\) and the respective marginal probabilities \(p(x)\) and \(p(y)\) based on the Fraser algorithm \[19\].

If a feature in the iris texture is uniformly distributed across different classes, its \(MI\) with respect to these classes is zero. If a feature is strongly differentially expressed for different classes, it should have large \(MI\). Thus, we use \(MI\) as a measure of relevance of features present in the iris with minimum redundancy among them.

In this paper, we propose to improve on previous approaches to using \(MI\) in feature selection \[20\], \[15\], \[21\] by computing the weight for each feature in relation to the other features. This helps to guide the selection method in search of the most relevant features.

C. Weighted Minimum redundancy and maximal relevance (W-mRMR)

Two forms of combining relevance and redundancy operations are used in \[20\], mutual information difference (MID), and mutual information quotient (MIQ). Thus the mRMR feature set is obtained by optimizing \(MID\) and \(MIQ\) simultaneously. Optimization of both conditions requires combining them into a single criterion function \[20\] as:

\[f^{mRMR}(fi) = MI(c; fi) - \frac{1}{|S|} \sum_{i = 1}^{n} MI(fi; fs)\]  

(2)

Figure 1: Representation of the 20 bands over the iris image.

We use the iris coded resulting from a particular instance of Gabor filters corresponding to the default values used in the IrisBee Software \[12\]. The bits selected by the feature selection methods represent this encoding information of the texture on the iris. If we make a drastic change to the parameters of the Gabor filter, then the features to select from would be different and so different features would be selected. Some Gabor filter parameters may be better or worse for representing the gender information in the iris texture. However, since a number of researchers have reported gender prediction results using various different filters \[12\], gender information can be captured by a variety of different filters.

The iris image is represented as an array of complex numbers with 20 rows and 240 columns. We extracted only the real component at this stage, to be used as the texture features. (The imaginary component is in a sense redundant \[13\].) Thus the binary iris code generated by IrisBee and used in this work is \(20 \times 240 = 4,800\) bits in size.

The IrisBee implementation incorporates a ‘fragile bits’ masking step \[7\]. For a given iris image, a bit in its corresponding iris code is defined as fragile if there is any substantial probability of it ending up a 0 for some images of the iris and a 1 for other images of the same iris. We used the IrisBee default threshold of masking the 25% most fragile bits for each iris code. Therefore, although the iris code is 4,800 bits as originally computed, fragile-bit masking reduces this to 3,600 potentially used for matching iris codes.

Figure 2 illustrates the creation of the iris code. Part (a) shows two example images, the left and right eyes of a female subject. Part (b) illustrates the image segmentation found by IrisBee. Part (c) shows the corresponding normalized images.

Part (d) illustrates the binary encoding of the texture filter result, the iris code. In the binary iris code, one concentric band of the iris is represented by 240 bits. Therefore, we have 20 bands with 240 bits each. Band 1 is the closest to the pupil and band 20 is closest to the sclera. (See Figure 1.) In section 4, we consider whether the texture features relevant to gender-from-iris are localized to a particular band of the iris.

B. Mutual Information

A feature is statistically relevant if its removal from a feature set will reduce the prediction power \[14\]. A feature \(fi\), can become redundant due to the existence of other relevant features, which provide similar prediction power as \(fi\). Some researchers propose to remove redundant features from the feature list, as this may improve the prediction accuracy \[15\]. Other researchers noticed that the removal of the redundant features may cause the exclusion of potential relevant features \[14\]. Therefore, they propose to find surrogate features by measuring feature correlations, or group features with similar patterns into feature clusters. Thus, we use Mutual Information to measure the relationship between features.
Figure 2: Images of the right and left eye of a female subject. (a) Original image, (b) Segmented image, with eyelids and eyelashes detection in yellow (c) Normalized image with mask in yellow (d) Binary iris code, with mask and fragile bits in white.

where, \( MI(c; f_i) \) measures the relevance of the feature \( f_i \) to be added for the class \( c \) and the term \( \sum_{f_s \in S} MI(f_i; f_s) \) estimates the redundancy of the \( f_{i\text{th}} \) feature with respect to the subset \( S \) of previously selected features \( f_s \).

However, there are two limitations for the above MI feature selection methods. Firstly, for any individual feature to be considered relevant it must be dependent with the target class. Groups of features that become relevant only as a group will not be discovered. The second weakness is that most of the methods simply consider pairwise feature dependencies. To overcome the above problem we propose a weighted approach. We propose to compute a weight for each feature in relation to the other features in the subset based on Euclidean distance and \( MI \), determining the relationships among features, and use the weight to guide the selection of the most relevant group of features. The procedure for computing the weights is based on the principles of the Relief-F method [22].

The weights are estimated as follows. To reduce computation, a given number of images are randomly sampled from the training data. For each randomly-sampled image, its \( k \) nearest neighbors from the same class (‘hits’) and \( k \) nearest neighbors from the opposite class (‘misses’) are found. See Figure 3. (Datasets with more than two classes are handled by finding the nearest neighbors from each class that are different from the currently sampled image). The nearest neighbors’ contribution is determined by the prior probability and \( MI \) of each class estimated from the training data.

The weight vector value is the average over the randomly-sampled images of the magnitude of the difference in distance to the \( k \) nearest hits and the distance to the \( k \) nearest misses, projecting on the \( R_i \) image. The greater the value of the weight, the greater the ability to differentiate between classes. The \( k \) nearest neighbors for the weight vector are selected from the complete set of training data. Therefore we know the position, the value and the \( k \) neighbors used for each feature. Then, we re-sort the weights and select the best subset using the Haussdorf distance [23] between them. We use these weights in the traditional feature selection methods but including the rank and weight for the best number of features, computed in the previous stage. Finally, using a wrapper method with SVM, we estimate the best number of features in relation with the \( k \) neighbors and classification rate employing forward elimination [24]. This approach allows us
to determine the most useful subset of the available features. This relationship was used as a measure of information, or as a measure of class separability based on that feature subset, in order to apply to W-mRMR and W-CMIM. The proposed feature selection method selects groups of features instead of pairs of features as done by MI traditional methods. Our weighting approach calculates the relevance and redundancy and adds the weight information function. Therefore $w_i(f_i)$ is used instead of the feature $f_i$, directly while computing MI score to make the ranking of relevance, $w_{mMR}(f_i) = \sum MI(f_i,f_s) w(f_i) w(f_s)$, where $w(f_i)^2$, $w(f_i)$ and $w(f_s)$ are used to weight the importance of the relation among features.

III. DATASET

We present results for three different iris image datasets. Examples are shown in Figure 2. For comparison to previous work, we present results on the dataset used in [6]. The dataset used in [6] was intended to have images from 1,500 different persons, but due to an error in image selection it contains images from less than 500 different persons. This means that results computed with this dataset in [6] are not person-disjoint in training and testing. For a more realistic estimate of accuracy, we also present results on a gender-from-iris (GFI) dataset that does represent 1,500 distinct persons and so does support person-disjoint training and testing. We also used an additional dataset for a validation test, we named UND_V. This dataset is person-disjoint from the original dataset with and additional 972 distinct persons and allows further validation of the results obtained on GFI dataset. Both the original dataset and this additional validation dataset will be made available at the URL referred to below.

Each dataset (GFI and corrected GFI) contains 3,000 images: 750 left-iris images from men, 750 right-iris images from men, 750 left-iris images from women and 750 right-iris images from women. Of the 1,500 distinct persons in the GFI dataset, visual inspection of the images indicates that about 1/4 are wearing clear contact lenses. Results computed for a person-disjoint train and test are generally expected to be lower than for train and test that is not person-disjoint. However, person-disjoint train and test should yield an accuracy estimate more realistic for generalization to new data. Comparing the results between these two datasets allows us to gauge the importance of using a person-disjoint train and test method.

A UND_V dataset contains 1,944 images: three left eye images and three right eye images for each of 175 males and 149 females. It is known that some subjects are wearing clear contact lenses, and evidence of this is visible in some images. Also, a few subjects are wearing cosmetic contact lenses in some images.

A. Is Gender Information Distributed Across All Bands?

In many irises, the bands nearer to the pupil have a more visually apparent texture pattern. Due to the structure of the eye, the bands nearer to the sclera are thought to change relatively less with changes in pupil dilation. The band of the
iris located near the pupil and the band located near the sclera will naturally have noise from inaccuracies in segmenting the iris boundary, and the boundary not being perfectly circular. Also, some previous researchers have suggested that some bands of the iris may be more important than others in terms of information for recognition of identity. Some bands perform better than others, because they are located further from the areas of occlusion or are more stable across different subjects, in terms of location and occlusion. For these various reasons, it seems worth considering whether useful gender-from-iris information is distributed across all bands of the iris.

To look at this question, a separate SVM was trained for each of the 20 bands of each of the left and the right iris. (Fragile bit information, segmentation mask, and feature selection are not used in this experiment). Figure 5 shows the gender classification accuracy using each one of the bands. Note that the accuracy from each of the 20 bands, for both the left and the right iris, is greater than 50%. This indicates that useful information for gender classification is present in all the bands of the iris. The accuracy appears to be generally higher in the inner half of the iris, bands 1 to 10, than in the outer half of the iris, bands 11 to 20. For the left iris, the best results are from bands 2, 4, 5 and 6 with 68.33%, 65.00%, 62.00% and 63.66% respectively. For the right iris, the best results are from bands 2, 6, 7 and 8 with 65.66%, 65.00%, 65.33% and 63.00% respectively. These results show that with only one band of an iris, corresponding to only 1/20 of the iris area, gender classification can be accomplished with the average accuracy as high as around 65%.

The analogous approach used with the non-subject-disjoint dataset in previous work [6] achieved accuracy of 87.33% for the left iris and 84.66% for the right iris. This shows how important it is to use a subject-disjoint dataset in order to obtain realistic accuracy estimates. The next four rows of Table I give the accuracy for the mRMR and the CMIM approaches to feature selection, with and without weighting of features. The number of bits selected from the iris code is given in parentheses below the accuracy number. We select the number of best features employing forward selection in steps of 100 features. Comparing the CMIM and mRMR approaches, we see that the CMIM achieves slightly higher accuracy than mRMR. This is true when feature weighting is used for both approaches, or not used for both. We also see that for either approach, weighting the features improves accuracy. Comparing the subject-disjoint evaluation results of this paper with the non-subject-disjoint results in [6], we see that non-subject-disjoint evaluation consistently results in unrealistically high accuracy estimates. For the highest-accuracy approach in Table I, W-CMIM, the optimistic bias introduced by non-subject-disjoint evaluation is about five percentage points. Finally, note that using W-CMIM to select features results in an improvement of about ten percentage points over using with whole iris.

W-CMIM uses conditional mutual information to measure the values of relevance and redundancy using the informa-
tion between two features previously selected and a third feature candidate. The W-mRMR approach uses traditional mutual information, making a trade-off between relevance and redundancy using only pairs of features. Thus, both weighted feature selection methods proposed in this paper may choose different distinctive bits of the iris-codes because they use different measures to estimate the relationships among different features.

Figure 7 illustrates the results of the W-CMIM feature selection. The features were selected separately for the left eye and right eye. The best 200 features are show in the top two rows, and the best 1,000 features are shown in the bottom two rows. In the first and third rows, the features are overlaid onto an example male pair of iris codes, and in the second and fourth rows, the features are overlaid onto an example male pair of iris codes. The selected features are distributed throughout the iris region, but there is some denser selection in the region that may correspond to where eyelid and eyelashes can occlude the iris texture.

C. Fusion of information from the left and right eyes

We consider multiple different approaches to fusing results from the left and the right iris. In one approach, feature-level fusion is done by concatenating the vectors of the left (1x4,800) and right iris (1x4,800) to obtain a person-level vector of 1x9,600 bits. Results of this approach are summarized in Table II. The results are largely analogous to those obtained in Table I for considering the left and the right iris individually. The W-CMIM approach to selecting features from the two-iris feature vector results in the best accuracy. However, the accuracy achieved is not significantly better than the accuracy achieved for a single iris. This could be due in part to the high dimensionality of the two-iris feature vector making the feature selection process more difficult.

In the second approach, we fuse results from the best features selected independently from the left and right irises based on results in Table I. These results are summarized in Table III. Analogous to the results in Tables I and II, the W-CMIM feature selection results in the highest accuracy. The W-CMIM fusion result in the last row of Table III is for fusion of the vector of best features from the left eye (1x2,200) and the right eye (1x3,200) as found in Table I. Using this 1x5,400 bit feature vector for the two irises achieves an accuracy of 89.00%. This represents a substantial improvement, of about five percentage points, over the accuracy obtained with a single iris.

We added two new experiments using the mirroring images from the left and the right iris-code. First, we select the best \( N \) features separately. The results show that the selected features were the same and only changed the feature index (relative position); therefore, the accuracy of the gender classification did not change. Second, we added another test with the fusion of the mirrored images. We used left images and its mirroring images for the same subject and the right images and its mirroring images. Then, we selected features from the fusion of 9,600 bits. Interestingly, the feature selection methods only select features from one side of the image (left or right) and removed the similar index features. This is because the method detects the redundancy between the features (bits) with equal mutual information values, thus mirrored features are
discharged and ranked in the last position.

D. Validation Test

The analysis of the misclassified images for the new validation database UND_V is presented in Table IV. The UND_V database does not have labels regarding problems for segmentation. Therefore, we assigned labels manually for the following tags: Straight eyelashes, Contact lenses (Texture-Cosmetic, Circular and clear contact lenses), Wrong segmentation, Blurred images and Well Segmented. Each image was labeled according to the main feature detected visually. For example, if one image has texture contact lenses and also has some blur, it is labeled with texture contact lenses tag because the texture contact lens impedes feature extraction for gender classification. For the case of fusion left and right iris codes, we use the tag with the main problem assigned to any of these images. For example if at least one of images (left or right iris) is tagged ‘Wrong segmentation’, the fusion is also tagged as ‘Wrong segmentation’.

In analyzing the results of the validation dataset, we found that mascara applied to eyelashes presents a problem for correct classification of female iris images. This is because the iris segmentation algorithm does not effectively mask out the instances of occlusion by eyelashes. For eyelashes without mascara, the effect on the image is relatively small, but with the presence of mascara it is larger. Of the images that have 'straight eyelashes' that cause occlusion of the iris texture, 67 of 117 left iris images were incorrectly classified, and 78 of 117 right iris images were incorrectly classified. An example of two of these misclassified images appears in Figure 8.

From Table IV, it can be observed that we have an important number of images labeled with wrong segmentation, due to the eyelashes (mascara), problems. Images with segmentation problems have lower gender classification accuracy. Therefore a comparison of the gender classification accuracy was made for manual segmentation and automatic segmentation with IrisBee software for the GFI Database and the UND_V database. The results were reported in Table V.

Table IV: Manual labelling of the UND_V database; Miss represent the misclassified images. Left and right Contact Lenses represents: 13 texture, 57 circular and 101 clear contact lenses.

<table>
<thead>
<tr>
<th>Labels</th>
<th>Left</th>
<th>Miss</th>
<th>%</th>
<th>Right</th>
<th>Miss</th>
<th>%</th>
<th>Fusion</th>
<th>Miss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Straight Eyelashes</td>
<td>117</td>
<td>67</td>
<td>57.26</td>
<td>117</td>
<td>78</td>
<td>66.67</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>Contact Lenses</td>
<td>171</td>
<td>47</td>
<td>27.49</td>
<td>171</td>
<td>54</td>
<td>31.58</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>Wrong Segmentation</td>
<td>399</td>
<td>60</td>
<td>15.04</td>
<td>418</td>
<td>11</td>
<td>16.75</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>Blurred images</td>
<td>3</td>
<td>2</td>
<td>66.67</td>
<td>24</td>
<td>35</td>
<td>37.50</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Well Segmented</td>
<td>282</td>
<td>22</td>
<td>7.80</td>
<td>242</td>
<td>8</td>
<td>0</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Subtotal</td>
<td>690</td>
<td>198</td>
<td>–</td>
<td>730</td>
<td>70</td>
<td>–</td>
<td>172</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>972</td>
<td>972</td>
<td>–</td>
<td>972</td>
<td>9</td>
<td>–</td>
<td>172</td>
<td></td>
</tr>
</tbody>
</table>

Figure 8: Segmentation results: (a) an image with straight eyelashes and with the occlusion mask, (b) an image with mascara and a wrong segmentation.

E. Statistical Test

The ANOVA test [26] was applied to determine whether or not differences among results were statistically significant. We applied this method to the results of different bands selected from the iris and to the results of left and right eyes. A small p-value ($p \leq 0.05$) indicates strong evidence against the null
Table V: Best gender classification rates for automatic segmentation in the GFI and UND_V database and for the manual segmentation for the UND_V database.

<table>
<thead>
<tr>
<th>Method</th>
<th>GFI_Dataset (%)</th>
<th>UND_V (%)</th>
<th>UND_V (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table I</td>
<td>85.33</td>
<td>76.66</td>
<td>78.33</td>
</tr>
<tr>
<td>Left - Right</td>
<td>+/- 0.89</td>
<td>+/- 0.95</td>
<td>+/- 1.05</td>
</tr>
<tr>
<td>Table II</td>
<td>Fusion</td>
<td>84.33 +/- 0.78</td>
<td>77.99 +/- 0.84</td>
</tr>
<tr>
<td>Left - Right</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table III</td>
<td>Best Left</td>
<td>89.00 +/- 0.86</td>
<td>82.33 +/- 0.90</td>
</tr>
<tr>
<td>Best right</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table VI shows the p-values from each left-right pair of bands. Bands 1, 18, 19, 20 have p value larger than 0.05 (p > 0.05), thus the information present on these bands is not statistically significantly different between the left and right bands. The p values for the other bands are lower than (p < 0.05) which means that there is different information from the bands of the left and right irises. Also we applied the ‘Bonferroni Adjustment’ [27] to the ANOVA test and reduced the threshold using p-value/number of test (0.05/10)=0.005. Due to the low p value shown in Table VI, the statistical significance of the bands did not change. The results are shown in Figure 9.

For the analysis of the results from the left and right eyes, we obtained a p value of 0.67, indicating that the accuracy is not statistically significantly different between the left and the right irises. We can see the results in Figure 10.

V. CONCLUSIONS

The major contributions of this work are as follows. One, we present results that achieve the highest accuracy yet published for person-disjoint gender-from-iris. Two, we present results that show that non-person-disjoint evaluation is biased toward unrealistically high accuracy. Three, this is the first paper to publish gender from iris-code. Four, feature selection is a broad field in continuous evolution, since selection of the most relevant and non-redundant features is not solved for complex problems such as in iris classification [14], [28], [29], [30]. Eliminating relevant or non-redundant features would result in poor behavior of the classifier. We present a new strategy to select the most important features (relevant and non-redundant) using neighbors’ information based on mutual information. Since relevant features are often unknown a priori, irrelevant and redundant features may be introduced to represent the domain. Our results show that improvements in gender classification are reached with our proposed feature selection method. We compared our results to those of traditional mutual information methods such as mRMR and CMIM. Besides, reducing the number of selected features reduces the computational burden in feature extraction. Five, we make two new iris image datasets available to the research community, so that others can study this problem, whereas previous works on this problem have generally not made their datasets available.

The binary iris code used for recognition is, in effect, a set of coarse texture features. Previous works in predicting gender from iris have relied on computing a separate, different texture representation [3], [4], [5]. Our results presented here are the first to show that gender can be successfully predicted from the binary iris code. There are clear computational advantages to predicting gender from the binary iris code rather than computing another different texture representation.

Our approach achieves gender prediction accuracy of 89% based on fusing the best features from the left and the right iris. The highest accuracy reported in previous work on gender-from-iris is 91% [6]. However, our results here are computed using a person-disjoint train-and-test, which produces an ac-
Table VI: ANOVA p-values with ‘Bonferroni Adjustment’ test for each left-right pair of bands.

<table>
<thead>
<tr>
<th>Band</th>
<th>Band 2</th>
<th>Band 3</th>
<th>Band 4</th>
<th>Band 5</th>
<th>Band 6</th>
<th>Band 7</th>
<th>Band 8</th>
<th>Band 9</th>
<th>Band 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.87</td>
<td>5.40e-9</td>
<td>1.16e-8</td>
<td>9.9e-14</td>
<td>6.28e-11</td>
<td>2.34e-6</td>
<td>1.94e-8</td>
<td>1.47e-5</td>
<td>2.46e-10</td>
<td>3.79e-10</td>
</tr>
<tr>
<td>2.37e-13</td>
<td>7.10e-11</td>
<td>2.45e-8</td>
<td>2.75e-10</td>
<td>8.66e-12</td>
<td>2.79e-10</td>
<td>4.04e-8</td>
<td>0.74</td>
<td>0.81</td>
<td>0.95</td>
</tr>
</tbody>
</table>

VI. Acknowledgements

This research was funded by CONICYT, FONDECYT through grant No. 1120613, and by the Department of Electrical Engineering, Universidad de Chile.

Thanks to Vince Thomas, Mike Batanian, Steve Lagree and Yingjie Gu for the work they have previously done in this research topic.

REFERENCES


