Gender Classification From the Same Iris Code Used for Recognition

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

Abstract—Previous researchers have explored various approaches for predicting the gender of a person based on the 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 the 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 the 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 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]–[6]. Most gender classification methods reported in the literature use texture features from the whole iris for classification or identification [3]–[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 and Bowyer [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 40×240 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 three, 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. Bowyer and Hollingsworth [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.
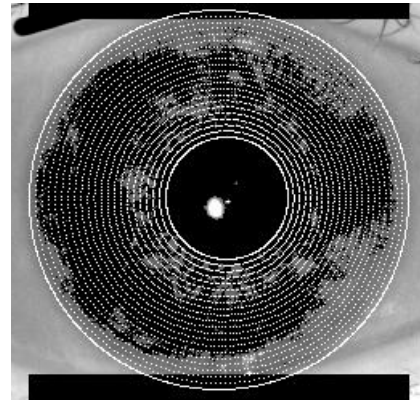

Fig. 1.   Representation of the 20 bands over the iris image.

We propose a new method to select a subset of bits from 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 II reviews background concepts in iris recognition and information theory, and explains our proposed method. Section III details the dataset used in our experiments. Section IV describes the experiments and results. Finally, conclusions and discussion are given in Section V.

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.
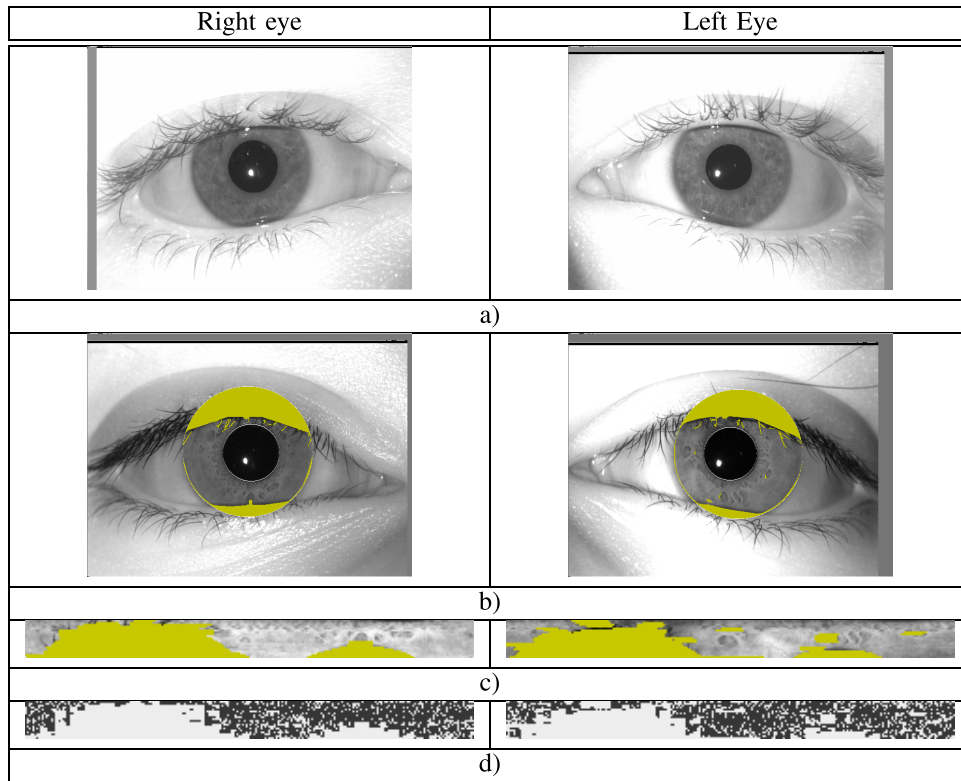
Fig. 2.   Images of the right and left eye of a female subject. (a) Original image. (b) Segmented image, with eyelids and eyelashes detection in yellow. (c) Normalized image with mask in yellow. (d) Binary iris code, with mask and fragile bits in white.

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]–[11], and therefore the possibility to classify gender from iris is a reality.

The radial resolution (r) and angular resolution (θ) 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 followings parameters: the wavelength (in pixels) of the log Gabor filter was 18 pixels and the ratio (σ/f_0) of the log Gabor filter was 0.5, where σ 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].

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.

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 is a measure of statistical independence that has two main properties. First, it can measure any kind of relationship between random variables, including nonlinear relationships [16]. Second, MI is invariant under transformations in the feature space that are invertible and differentiable, e.g., translations, rotations, and any transformation preserving the order of the original elements of the feature vectors [18]

The MI between two variables, x and y, is defined based on their joint probability distribution $p(x; y)$ and the respective marginal probabilities $p(x)$ and $p(y)$ as:

$$MI(x, y) = \int \int p(x, y) log \frac{p(x, y)}{p(x)p(y)} dx dy \qquad (1)$$

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 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 [15], [20], [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.
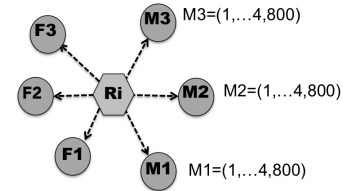

Fig. 3. Representation of modified Relief-F with three nearest neighbors, k=3. Ri is a random image; F represents Nearest Hits and M represent Nearest Misses. In this work during the normalization process determines the size of the rectangular iris image. We use a rectangular image of $20(r) \times 240(\theta) = 4,800$.

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:

$$f^{mRMR}(fi) = MI(c; fi) - \frac{1}{|S|} \sum MI(fi; fs) \qquad (2)$$

where, $MI(c; fi)$ measures the relevance of the feature f_i to be added for the class c and the term $\frac{1}{|S|} \sum_{fi \in S} MI(fi; fs)$ estimates the redundancy of the fi_{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 $w_i(fi)$ is used instead of the feature fi, directly while computing MI score to make the ranking of relevance, $f^{W-mRMR}(f_i) =$

$$MI(c\,;\,fi) \times w(fi)^2 - \frac{1}{|S|} \sum MI(fi\,;\,fs)$$
$$\times w(fi) \times w(fs), \qquad (3)$$

where $w(fi)^2$, $w(fi)$ and $w(fs)$ 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.[1]

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.

IV. EXPERIMENTS AND RESULTS

We present results for three experiments. In Experiment 1, we consider the accuracy of gender classification using individual bands of the iris, as illustrated in Figure 1. A band of the iris that contains no useful information for classifying gender should have an accuracy of approximately 50%. To validate the results we add a new validation dataset named UND_V to analyze how well the classifier generalizes to new data. In Experiment 2, we classify gender using feature selection based on MI as described in the Section II. We perform feature selection and estimate accuracy separately for the left and right iris because some systems work with a single iris. Also, even systems that can work with two irises need to be able to work with only one if the other is covered for some reason. In Experiment 3, we evaluate possible approaches to fusing information from the left and right iris.

In the second and third experiments we use a segmentation mask of the same size as the rectangular image and additionally masked by IrisBee default 25% of fragile bits [7]. There is a unique mask for each image, which means that a particular bit may be masked in one image and not in another image.

A training portion of the 1,500-person dataset was created by randomly selecting 80% of the males and 80% of the females. We used 10-fold cross-validation on this training set of 80% of the original data to select parameters of each method. Once parameter selection is finalized, the selected parameterization of the method is trained on the full 80% training data, and a single evaluation is made on the 20% test data. In each experiment, an SVM classifier with Gaussian kernel was trained using LIBSVM implementation [25]. To validate the results we add a new validation dataset named UND_V to analyze how well the classifier generalizes to new data.

It is important to note that iris images of different persons, or even the left and right iris images for a given person, may not present exactly the same imaging conditions. The illumination by the LEDs may come from either side of the sensor, specular highlights may be present in different places in the image, the inclination of the head may be different, the eyelid occlusion may be different, and so forth. See Figure 4. Therefore we expect to select different features from each eye.

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
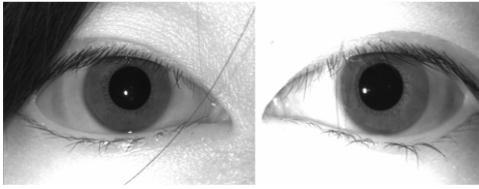
Fig. 4. Illustration of different imaging conditions for the two eyes of one subject.
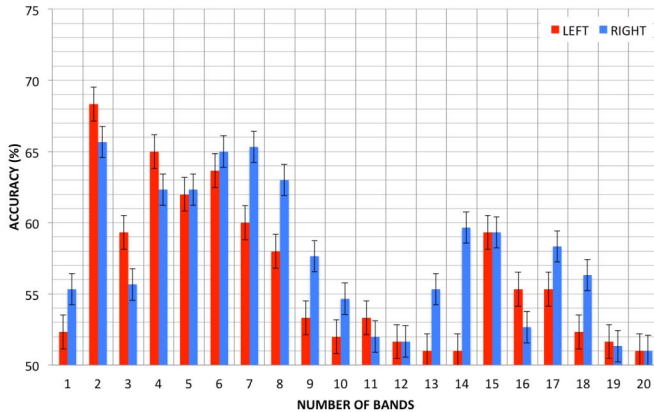

Fig. 5. Gender Classification rate using only one band at a time from GFI database. The classification rate for left-iris data is in red, and for right-iris data is in blue.

Fig. 6. Gender classification rate using only one band at time from UND_V database. The classification rate for the left iris data is in red, and for right iris data is in blue.

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%.

Figure 6 shows the classification rate by band of the iris computed for 1/3 of the UND_V database described
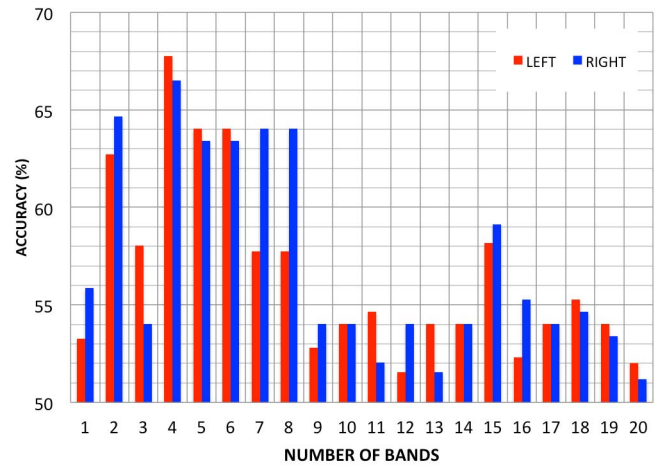
in Section III, one left eye image and one right eye for each of 175 males and 149 females, subject-disjoint from the GFI dataset, created a person disjoint dataset. The results are similar to those in Figure 5, validating that the results in Figure 5 are not dependent on the particular persons in the GFI dataset, and that useful information for gender classification is present in all the bands.

B. Feature Selection From the Bands

The results of our first experiment show that there is gender-from-iris information in all bands of the iris. However, using the whole iris code would result in a large feature vector. In experiment two, we apply the feature selection methods discussed in Section II to find a reduced feature vector and compare the accuracy to that of using the whole iris code for gender prediction.

The top row of Table I shows the accuracy achieved when an SVM is trained for gender classification using the whole 1,200-element training set, where each training element is a 4,800-bit iris code. This approach results in 77.33% and 74.66% accuracy for the left iris and the right iris, respectively. 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],

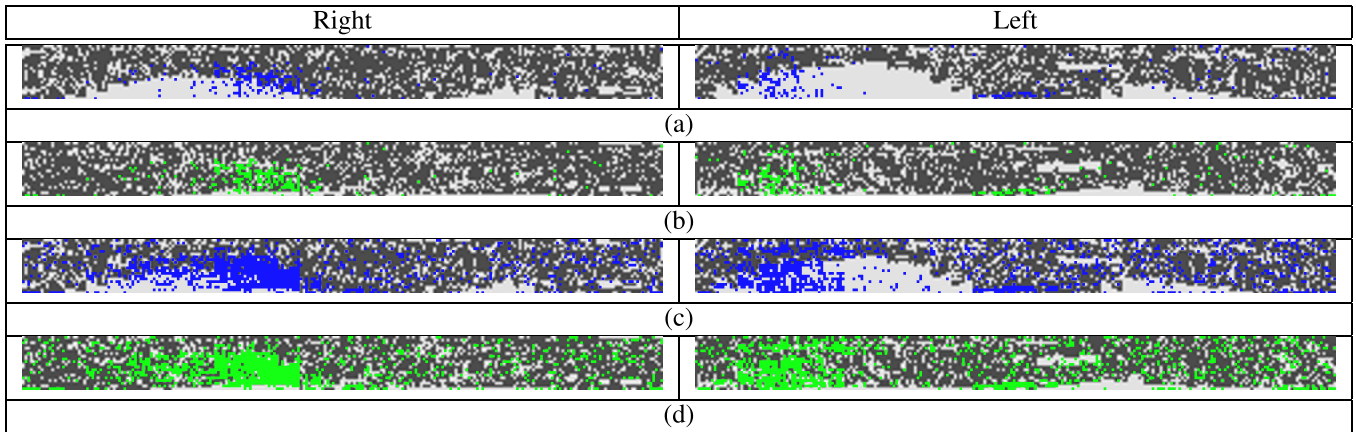| Right	Left

(a)

(b)

(c)

(d)

Fig. 7. Four different pairs of (left, right) iris images are shown. Features selected for female irises are shown in blue in a) and c), and for male irises are shown in green in b) and d). The best 200 features selected are shown in a) and b), and the best 1,000 features selected are shown in c) and d). The feature selections shown are those found with W-CMIM.

TABLE I

GENDER CLASSIFICATION RATES FOR THE LEFT AND RIGHT IRIS CODE CONSIDERING ALL BITS AND THOSE SELECTED BY MUTUAL INFORMATION MEASURES mRMR, CMIM, W-mRMR AND W-CMIM. IN PARENTHESIS APPEAR THE NUMBER OF SELECTED FEATURES. k IS THE NUMBER OF NEAREST NEIGHBORS

Method	Left eye (%)	Right eye (%)	Left eye (%)	Right eye (%)
	[6]	[6]	GFI	GFI
Whole Iris Code	87.33 +/- 0.72 (4,800)	84.66 +/- 0.77 (4,800)	77.33 +/- 0.78 (4,800)	74.66 +/- 0.75 (4,800)
mRMR	87.66 +/- 0.67 (2,800)	87.00 +/- 0.63 (3,400)	79.66 +/- 0.84 (900)	81.00 +/- 0.87 (1,100)
CMIM	88.66 +/- 0.66 (1,900)	88.66 +/- 0.71 (4,200)	82.66 +/- 0.89 (900)	81.66 +/- 0.91 (1,200)
W-mRMR	88.00 +/- 0.73 (900) (k=10)	85.66 +/- 0.70 (400) (k=10)	84.00 +/- 0.78 (900) (k=10)	82.00 +/- 0.75 (1,200) (k=10)
W-CMIM	**91.00 +/- 0.63** (1,200) (k=5)	**89.00 +/- 0.60** (3,200) (k=5)	**85.33 +/- 0.73** (2,200) (k=10)	**84.33 +/- 0.69** (3,200) (k=10)

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 information 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 female 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 ($1 \times 4,800$) and right iris ($1 \times 4,800$) to obtain a person-level vector of $1 \times 9,600$ 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 ($1 \times 2,200$) and the right eye ($1 \times 3,200$) as found in Table I. Using this $1 \times 5,400$ 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

TABLE II
GENDER CLASSIFICATION RATES FOR THE LEFT AND RIGHT IRIS CODE.
IN PARENTHESIS APPEAR THE NUMBER OF SELECTED FEATURES.
k IS THE NUMBER OF NEAREST NEIGHBORS

Method	Accuracy (%)	Male (%)	Female (%)
Raw Fusion	72.66 +/- 0.91 (9,600)	73.66	71.66
mRMR Fusion	82.16 +/- 0.90 (4,000)	82.33	81.99
CMIM Fusion	83.00 +/- 0.90 (4,000)	83.66	82.33
W-mRMR Fusion	82.33 +/- 0.87 (4,000) (k=5)	83.00	81.66
W-CMIM Fusion	**84.16 +/- 0.92** (5,000) (k=5)	**84.99**	**83.33**

TABLE III
GENDER CLASSIFICATION RATES FOR THE FUSION OF LEFT AND RIGHT
IRIS CODES USING GFI DATABASE. IN PARENTHESIS APPEAR THE
NUMBER OF SELECTED FEATURES. k IS THE NUMBER
OF NEAREST NEIGHBORS

Method	Accuracy (%)	Male (%)	Female (%)
Best- mRMR Fusion	82.00 +/- 0.75 (2,000)	83.00	81.00
Best -CMIM Fusion	85.00 +/- 0.77 (2,100)	84.66	85.33
Best W-mRMR Fusion	85.33 +/- 0.70 (2,100)(k=5)	85.00	86.66
Best W-CMIM Fusion	**89.00 +/- 0.68** (5,400)(k=5)	88.33	89.66

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

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

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

	Automatic Segmentation		Automatic Segmentation		Manual Segmentation	
Best case	GFI_Dataset (%)		UND_V (%)		UND_V (%)	
Method	Left	Right	Left	Right	Left	Right
Table I Left - Right	85.33 +/- 0.89	84.33 +/- 0.95	76.66 +/- 1.05	78.33 +/- 0.87	84.66 +/- 0.96	85.33 +/- 0.91
Table II Fusion Left - Right	84.33 +/- 0.78		77.99 +/- 0.84		81.33 +/- 0.90	
Table III Best Left Best right	89.00 +/- 0.86		82.33 +/- 0.90		85.99 +/- 0.89	

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 textured 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.

TABLE VI

ANOVA p-VALUES WITH 'BONFERRONI ADJUSTMENT' TEST FOR EACH LEFT-RIGHT PAIR OF BANDS

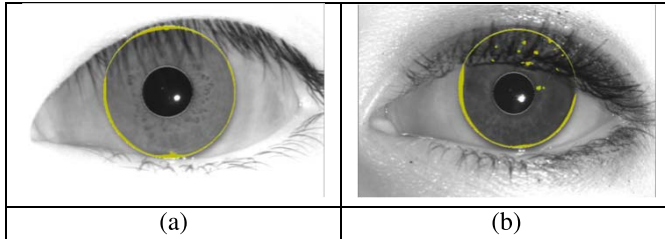| Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7	Band 8	Band 9	Band 10
0.87	5.40e-9	1,16e-8	9.9e-14	6.28e-11	2.34e-6	1.94e-8	1.47e-5	2.46e-10	3.79e-10
Band 11	Band 12	Band 13	Band 14	Band 15	Band 16	Band 17	Band 18	Band 19	Band 20
2.37e-13	7.10e-11	2.45e-8	2.75e-10	8.66e-12	2.79e-10	4.04e-8	0.74	0.81	0.95

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

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 Iris-Bee software for the GFI Database and the UND_V database. The results were reported in Table V.

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 hypothesis, so we can reject the null hypothesis. A large p-value ($p > 0.05$) indicates weak evidence against the null hypothesis, so we fail to reject the null hypothesis. P values are calculated based on the assumptions that the null hypothesis is true. For the results of different band numbers of the iris, we get a p-value of 1.5580e-05, suggesting that the hypothesis that the information is the same across the bands can be rejected. The results are shown in Figure 9.

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.

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
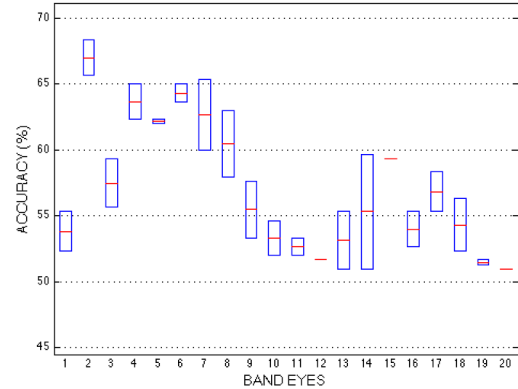

Fig. 9. Mean accuracy and standard deviation for different bands of the left and right eyes. The central mark in red is the mean; the edges of the box are the 25th and 75th percentiles.
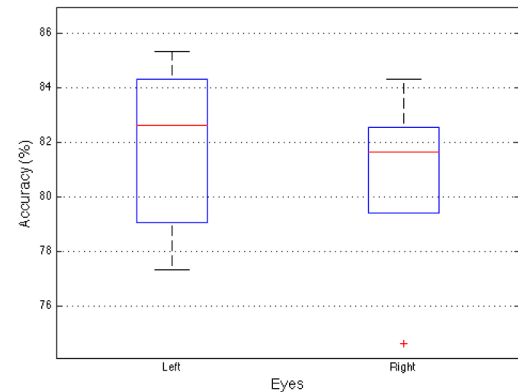

Fig. 10. Mean accuracy and standard deviation for the left and right eyes. The central mark in red is the mean; the edges of the box are the 25th and 75th percentiles.

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]–[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]–[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 accuracy estimate that generalizes better to unseen data. In our comparison of results that are person-disjoint and non-person-disjoint manner, non-person disjoint accuracy estimates came out 5% or more higher than person-disjoint estimates, for the same methodology and same size dataset. (See Table I.)

Our initial experiments suggest that information relevant to gender-from-iris is distributed throughout the whole iris region. (See section 4.A.) However, our results also show that selecting features that represent the areas of the iris that allow the most stable texture computations improves performance relative to using features from all areas of the iris. This is likely due to a combination of the high feature dimensionality when using the whole area of the iris, and the fact that segmentation inaccuracies can make some areas of the iris less stable for texture computation.

We explored feature selection using four mutual information measures: mRMR, CMIM using pairs of features and W-mRMR and W-CMIM. We found that using W-CMIM, for selecting groups of features, results reached the best accuracy. It obtained 85.33% and 84.33% on the left and right iris code experiments, respectively. It also obtained 89% using the best bits for the left and right iris code.

In this way, the problem dimensionality can be reduced, improving classification rate and shortening the computational time required. In the presence of a very large number of features (tens of thousands), it is common to find a large number of features that do not contribute to the classification process because they are irrelevant or redundant with respect to a particular class.

Certainly one goal of this research is to establish the relationship between the features useful for predict gender and biological insight. The results reported in this paper add to the evidence that it is possible to classify gender from iris texture; in this paper, even the highly abstract texture representation that is the iris code. We also show that gender information is distributed across different bands of the iris. The statistical analysis supports this assumption due to the mean of the bands are different. The results also show that the difference in gender-from-iris accuracy between left and right eyes is not statistically significant. It appears to be due to only the random variation of dataset and the influence of the illumination, segmentation quality, and other factors. However the same features selected (bits from the iris-code) from GFI database performed well in a subject-disjoint dataset such as UND_V. This shows that the gender information is available in the structure of the iris. The authors believe that more work must be done to establish general conclusions in relation to biological insight and features selected, such as: test with different sensor types and other datasets.

REFERENCES

[1] J. E. Tapia and C. A. Perez, "Gender classification based on fusion of different spatial scale features selected by mutual information from histogram of LBP, intensity, and shape," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 3, pp. 488–499, Mar. 2013.

[2] J. Daugman, "How iris recognition works," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 21–30, Jan. 2004.

[3] A. Bansal, R. Agarwal, and R. K. Sharma, "SVM based gender classification using iris images," in *Proc. 4th Int. Conf. Comput. Intell. Commun. Netw. (CICN)*, Nov. 2012, pp. 425–429.

[4] S. Lagree and K. W. Bowyer, "Predicting ethnicity and gender from iris texture," in *Proc. IEEE Int. Conf. Technol. Homeland Secur. (HST)*, Nov. 2011, pp. 440–445.

[5] V. Thomas, N. V. Chawla, K. W. Bowyer, and P. J. Flynn, "Learning to predict gender from iris images," in *Proc. 1st IEEE Int. Conf. Biometrics, Theory, Appl., Syst. (BTAS)*, Sep. 2007, pp. 1–5.

[6] J. E. Tapia, C. A. Perez, and K. W. Bowyer, "Gender classification from iris images using fusion of uniform local binary patterns," in *Proc. Eur. Conf. Comput. Vis., Soft Biometrics Workshop (ECCV)*, 2014, pp. 751–763.

[7] K. P. Hollingsworth, K. W. Bowyer, and P. J. Flynn, "The best bits in an iris code," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 6, pp. 964–973, Jun. 2009.

[8] K. W. Bowyer, K. Hollingsworth, and P. J. Flynn, "Image understanding for iris biometrics: A survey," *Comput. Vis. Image Understand.*, vol. 110, no. 2, pp. 281–307, 2008.

[9] J. Daugman, "Iris recognition at airports and border-crossings," in *Encyclopedia of Biometrics*, S. Z. Li and A. Jain, Eds. Boston MA, USA: Springer, 2009, pp. 819–825.

[10] *Canadian Border Services Agency*, CANPASS, 1996. [Online]. Available: http://www.cbsa-asfc.gc.ca/prog/canpass/canpassair-eng.html

[11] *Unique Identification Authority of India*, UIDAI, 2014. [Online]. Available: https://uidai.gov.in/about-uidai.html

[12] X. Liu, K. W. Bowyer, and P. J. Flynn, "Experiments with an improved iris segmentation algorithm," in *Proc. 4th IEEE Workshop Automat. Identificat. Adv. Technol.*, Oct. 2005, pp. 118–123.

[13] J. Daugman, "Information theory and the iris code," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 2, pp. 400–409, Feb. 2016.

[14] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, Eds., *Feature Extraction: Foundations and Applications* (Studies in Fuzziness and Soft Computing). Secaucus, NJ, USA: Springer-Verlag, 2006.

[15] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," in *Proc. IEEE Bioinformatics Conf. (CSB)*, Aug. 2003, pp. 523–528.

[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 1991.

[17] P. A. Estévez, M. Tesmer, C. A. Perez, and J. M. Zurada, "Normalized mutual information feature selection," *IEEE Trans. Neural Netw.*, vol. 20, no. 2, pp. 189–201, Feb. 2009.

[18] S. Kullback, *Information Theory and Statistics*. New York, NY, USA: Dover, 1997.

[19] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Phys. Rev. A*, vol. 33, no. 2, pp. 1134–1140, Feb. 1986.

[20] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[21] F. Fleuret, "Fast binary feature selection with conditional mutual information," *J. Mach. Learn. Res.*, vol. 5, pp. 1531–1555, Dec. 2004.

[22] I. Kononenko, "Estimating attributes: Analysis and extensions of RELIEF," in *Proc. 7th Eur. Conf. Mach. Learning*, Catania, Italy. Berlin, Germany: Springer, 1994, pp. 171–182.

[23] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 9, pp. 850–863, Sep. 1993.

[24] J. M. Cadenas, M. C. Garrido, and R. MartíNez, "Feature subset selection filter-wrapper based on low quality data," *Expert Syst. Appl.*, vol. 40, no. 16, pp. 6241–6252, 2013.

[25] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.

[26] A. Cuevas, M. Febrero, and R. Fraiman, "An anova test for functional data," *Comput. Statist. Data Anal.*, vol. 47, no. 1, pp. 111–122, 2004.

[27] P. Sedgwick, "Statistical question multiple significance tests: The Bonferroni correction," in *Proc. BMJ*, vol. 344. Jan. 2012, p. 170.

[28] J. R. Vergara and P. A. Estévez, "A review of feature selection methods based on mutual information," *Neural Comput. Appl.*, vol. 24, no. 1, pp. 175–186, Jan. 2014.

[29] H.-H. Hsu, C.-W. Hsieh, and M.-D. Lu, "Hybrid feature selection by combining filters and wrappers," *Expert Syst. Appl.*, vol. 38, no. 7, pp. 8144–8150, 2011.

[30] P. L. Williams and R. D. Beer, "Nonnegative decomposition of multivariate information," *Comput. Res. Repository (CoRR)*, vol. 1004, p. 2515, 2010. [Online]. Available: http://arxiv.org/abs/1004.2515

Claudio A. Perez (SM'04) received the B.S. degree and the P.E. degree in electrical engineering and the M.S. degree in biomedical engineering from the Universidad de Chile, in 1980 and 1985, respectively, and received the Ph.D. degree from Ohio State University, in 1991. He was a Fulbright Student with Ohio State University, where he obtained a Presidential Fellow in 1990. He was a Visiting Scholar with the University of California, Berkeley, in 2002, through the Alumni Initiatives Award Program from the Fulbright Foundation. He was the Department Chairman from 2003 to 2006 and is currently the Director of the Office of Academic and Research Affairs with the School of Engineering, Universidad de Chile. He is a Professor with the Department of Electrical Engineering, Universidad de Chile. His research interests include biometrics, image processing applications, and pattern recognition. He is a Senior Member of the Systems, Man and Cybernetics and IEEE CIS societies.

Kevin W. Bowyer is currently the Schubmehl-Prein Professor of Computer Science and Engineering with the University of Notre Dame, and serves as the Chair of the Department. His research interests range broadly over computer vision and pattern recognition, including biometrics and data mining. He is a fellow of the IAPR for the contributions to computer vision, pattern recognition, and biometrics, and a Golden Core Member of the IEEE Computer Society. He received a 2014 Technical Achievement Award from the IEEE Computer Society, with the citation "For pioneering contributions to the science and engineering of biometrics." He is serving as the Chair of the Research and Innovation Track at Biometrics 2016, and the General Chair of the 2017 IEEE International Conference on Automatic Face and Gesture Recognition; has served as the Editor-in-Chief of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE and the IEEE BIOMETRICS COMPENDIUM; and is currently serving on the Editorial Board of IEEE ACCESS. His most recent book is the *Handbook of Iris Recognition*, edited with Dr. M. Burge.

Juan E. Tapia received the P.E. degree in electronics engineering from Universidad Mayor, in 2004, the M.S. degree in electrical engineering from the Universidad de Chile, in 2012, and the Ph.D. degree from the Department of Electrical Engineering, Universidad de Chile, in 2016. His main interests are pattern recognition and machine learning applied to biometrics, gender classification, feature fusion, and feature selection.