IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 3, 2019, accepted February 25, 2019, date of publication March 1, 2019, date of current version March 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2902470

Gender Classification From NIR Images by Using Quadrature Encoding Filters of the Most Relevant Features

JUAN E. TAPIA[1], (Graduate Student Member, IEEE), AND CLAUDIO A. PEREZ[2,3], (Senior Member, IEEE)

[1]Universidad Tecnologica de Chile-INACAP, Santiago 8580704, Chile
[2]Advanced Mining Technology Center, Universidad de Chile, Santiago 8370451, Chile
[3]Department of Electrical Engineering, Universidad de Chile, Santiago 8370451, Chile

Corresponding author: Juan E. Tapia (j_tapiaf@inacap.cl)

This work was supported in part by the Fondecyt under Grant 1161034, Department of Electrical Engineering and the Advanced Mining Technology Center, Universidad de Chile, and in part by the Fondecyt Iniciacion through the Universidad Tecnologica de Chile- INACAP under Grant 11170189.

ABSTRACT In the past few years, accuracy in determining gender from iris images has increased significantly, approaching levels that make novel applications of this biometric technology feasible. In this paper, we report the gender classification rate by using a 2-D Quadrature Quaternionic filter, and a selection of the most relevant features from the normalized iris images. We encoded the phase information of the normalized images using 4 bits per pixel with a 2-D-Gabor filter and selected the best bits from the four resulting images (1 real and 3 imaginary) instead of the 1-D log-Gabor traditional encoding method. We used traditional handcrafted and automatic methods to select and extract the most relevant features from the whole iris images, blocks from images, and pixel features and compared how effective these methods were in separating features from female and male iris images. Selecting iris blocks and features reduce the computational time and, at a basic science level, is of great value in understanding what information features, as well as pixels from the iris, can be extracted to classify gender. The Quaternionic-Code with the complementary feature selection method achieved the best results on the GFI-UND database with 93.45% for the left iris and 95.45% for the right iris, both with 2400 selected features. We compared our results and found them to be advantageous to the best results previously published, and also to those obtained using convolutional neural network feature extraction.

INDEX TERMS Soft-biometrics, gender classification, gender from iris, iris relevant features.

I. INTRODUCTION

Soft-biometrics using iris information is a new form of estimating demographic information or forensic information, such as gender, ethnicity, age and, emotions [1], [19], [19], [27], [30], [32], [33]. In a biometric recognition system, gender information may lead to searching only half of the database [42]. 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. 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 [39]. Gender classification is also important for demographic information collection, marketing research, and real-time electronic marketing [22], [28], [38]. See Figure 1.

Most gender classification methods reported in the literature use all of the features extracted for classification purposes. In image understanding, raw input data often has very high dimensionality and a limited number of samples. In this area, feature selection plays an important role in improving accuracy, efficiency and scalability of the object identification process [46]. According to Bengio et al. [3], "The success of machine learning algorithms generally depends on

The associate editor coordinating the review of this manuscript and approving it for publication was Nilanjan Dey.

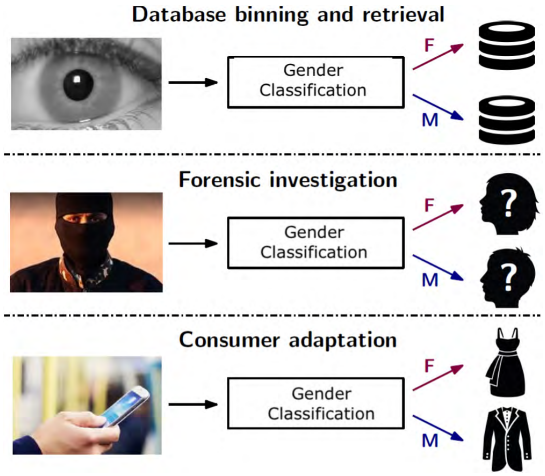

FIGURE 1. Different application scenarios where soft-biometrics such as gender classification from iris images can be applied. Images based on [8].

data representation, and we hypothesize that this is because different representations can entangle and hide more or less the different explanatory factors of variation behind the data''.

The statistical significance tests such as: t-test, ANOVA and Mutual Information (*MI*) have been used previously as a criteria for feature selection [21], [23], [39], [46], [47]. Nevertheless, only a few papers included statements or references to assumption about the variance homogeneity necessary for the application of the t-test, ANOVA and *MI*.

For instance, the t-test and ANOVA are fairly robust to moderate departures from its assumptions of normally-distributed data and equality of variance (homogeneity) except in the presence of very small or unequal sample sizes, which can considerably decrease the statistical power of the analyses.

The p-value, the average, the relevance and the redundancy can be used to identify the best number of features. Features that do not show clear patterns of differential expression are filtered out, by performing statistical group comparisons. However, if the size of the features has not been properly estimated before the statistical comparisons (e.g. ANOVA or t-test), then spurious predictions and errors can be seriously misleading. In fact, undetected significant differences may be explained by a lack of statistical power for detecting true differences between features or as a result of inadequate sample sizes.

This paper shows that only a few soft-biometrics research studies [39] discuss the quantity and the quality of sample size requirements in biometric experiments, which are fundamental factors to accomplish the validation of the statistical analyses.

Gender classification using iris information is a rather new topic, with only a few papers published [2], [14], [25], [44].

Thomas et al. [44] were the first to explore gender-from-iris, using normalized images acquired with an LG 2200 sensor. They segmented the iris region and employed machine learning techniques to develop models that predicted gender based on the iris texture features. They segmented the iris region, created a normalized iris image, and then 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 and were able to reach an accuracy close to 80%.

Lagree and Bowyer [25] experimented with normalized iris images acquired using an LG 4000 sensor. They 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 was not processed by the log-Gabor filters that were used by the IrisBEE software [29] to create the 'iris code' for biometric purposes and did not use any geometric features to develop models that predicted gender and ethnicity based on the iris texture feature. These are the differences from features computed by Thomas in [44]. This approach reached an accuracy close to 62% for gender and close to 80% for ethnicity.

Bansal et al. [2] experimented with normalized iris images acquired with a Cross Match SCAN-2 dual-iris camera. A statistical feature extraction technique based on the correlation between adjacent pixels was combined with a 2D wavelet tree based on feature extraction techniques to obtain significant features from the iris image. This approach reached an accuracy of 83.06% for gender classification. Nevertheless, the database used in this experiment was very small (300 images) compared to other studies published in the literature.

Tapia et al. [42] experimented with normalized iris images acquired using an LG 4000 sensor to classify the gender of a person based on the analysis of the iris texture features. We explored using different implementations of Local Binary Patterns (LBP) from the iris image using occlusion mask. It was shown that Uniform LBP with concatenated histograms significantly improved the accuracy of gender prediction relative to using the whole iris image. Results achieved over 91% correct gender prediction using a non-person-disjoint dataset and the texture of the iris.

Costa-Abreu et al. [14] explored the gender prediction task with respect to three approaches using only geometric features, only texture features and both geometric and texture feature extracted from normalized iris images. This work used a BioSecure Multimodal DataBase (BMDB) and these images were taken using a LG Iris Access EOU-3000. They were able to achieve over 89.74% correct gender prediction using the texture of the iris. Nevertheless, the dataset is not available to other researchers.

Bobeldyk and Ross [4] explored the gender prediction accuracy associated with 4 different regions from NIR iris images: the extended ocular region, the iris-excluded ocular region, the iris-only region, and the normalized iris-only region. They used a Binarized Statistical Image Feature (BSIF) texture operator to extract features from the regions previously defined. The ocular region reached the

best performance with 85.7 % while the normalized images exhibited the worst performance, with almost a 20 percent point of difference in performance over the ocular region (65%). It was concluded that the normalization process could be filtering out useful information.

Tapia et al. [39] predicted gender directly from the same binary iris-code images that could be used for recognition. It was found that information for gender prediction was distributed across the iris, rather than localized in particular concentric bands. It was also found that using selected features representing a subset of the iris region achieved a better accuracy was achieved compared to using features representing the whole iris region reaching 89% correct gender prediction using the fusion of the best features of iris-code from the left and the right eyes.

Kuehlkamp et al. [24] estimated the accuracy using a mean of N person-disjoint train and test partitions, and considering the effect of makeup. They also showed that classification based on the occlusion masks, disregarding completely the iris texture, results in an accuracy of approximately 60%. They also showed that simple averaging of the normalized iris image intensity and thresholding can result in approximately 60% gender-from-iris accuracy.

Only a few methods have used Deep Learning on gender classification such as gender with periocular and normalized NIR images respectively [24], [31], [41].

Tapia and Aravena [41] used a supervised and a semi-supervised method with Convolutional Neural Networks (CNN) to classify gender, reaching an accuracy of 83.00 % with a larger number of normalized images and data-augmentation.

The main contributions of this paper are the following: We measured the quality and quantity of the information present in the iris normalized-encoded images and explored the application of 2D Gabor filters in order to extract 2D spatial information using four pixels per bit, instead of the traditional 1D Gabor filter approach with 2 pixels per bit. In more detail, first, we evaluate the gender classification accuracy achieved using the whole normalized image of the iris with several machine learning classifiers. Then, we present a new strategy for selecting the most important features, relevant, and non-redundant, analyzing different fitness values using a GA, p-value, t-test, and MI. The quality of different blocks and features from the normalized iris image is estimated to improve gender classification. Second, we compare the performance of our method with automatic feature extraction using Deep Learning-bottleneck and Transfer Learning with VGG-19 [35]. Third, we encode the phase information of the normalized iris images using 2D quadrature filtering. A few attempts have been made to extend a 1D signal to two dimensions, e.g., the partial Analytics Signal, Total Analytics signal, Monogenic Signal, and Quaternionic signal [5], [11]. We evaluated the Quaternionic-Code presented in [11], which is essentially a binarized phase representation of an iris image using quaternion filters. The use of Quaternionic-Code is motivated by its ability to localize spatial and frequency domain phase information jointly in the normalized iris images. We compare our results to those previously published and summarized them in Table 1.

TABLE 1. Summary of gender classification using eyes. I represents: Iris Images, P represents: Periocular Images, CP represent: Cell-phones Images. The best results of this paper is highlighted in bold.

Paper	I/P	Source	N^o Subjects	Type	Acc. %
V .Thomas et al. [44]	I	Iris	N/A	NIR	75.00
S. Lagree et al. [25]	I	Iris	300	NIR	62.17
A. Bansal et al. [2]	I	Iris	200	NIR	83.60
J. Tapia et al. [42]	I	Iris	1,500	NIR	91.00
M. Fairhurst et al. [14]	I	Iris	200	NIR	89.74
J. Tapia et al. [39]	I	Iris	1,500	NIR	89.00
D. Bobeldyk et al. [4]	I/P	Iris	1,083	NIR	85.70
Kuehlkamp et al. [24]	I/P	Iris	1,500	NIR	80.00
J. Tapia. [40]	I/P	Iris	1,500	NIR	79.33
J. Tapia et al. [41]	I	Iris	1,500	NIR	83.00
J. Merkow et al. [27]	P	Faces	936	VIS	80.00
C. Chen et al. [13]	P	Faces	1,003	NIR/Thermal	93.59
Castrillon-Santana et al. [10]	P	Faces	1,500	VIS	92.46
Rattani et al. [33]	P	Iris	550	VIS/ CP	91.60
J. Tapia et al. [43]	P	Iris	120/120	NIR/VIS	90.00
This proposal	**I**	**Iris**	**1,500/1,500**	**NIR**	**95.45**

II. IRIS FEATURE EXTRACTION

The iris feature extraction process involves the following steps. ''First, a camera acquires an image of the eye. All commercial iris recognition systems use near-infrared illumination, to be able to image iris texture of both 'dark' and 'light' eyes'' [42]. Next, the iris region is located within the image. The annular region of the iris is transformed from raw image coordinates to normalized polar coordinates. This results in what is sometimes called an 'unwrapped' or 'rectangular' iris image. 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 [6]. Iris recognition systems operating on these principles are widely used in a variety of applications around the world [9], [17], [45].

The radial resolution (r) and angular resolution (θ) used during the normalization or 'unwrapping' step determine the size of the normalized iris image, and can significantly influence the iris recognition rate. This unwrapping is referred to as using Daugman's rubber sheet model [16]. In this work, we use a normalized image of 20 (r) $\times$ 240 (θ), created using Daugman's method and Osiris implementation [37]. Both implementations also create a segmentation mask of the same size as the normalized image. The segmentation mask indicates the portions of the normalized iris image that are not valid due to occlusion by eyelids, eyelashes or

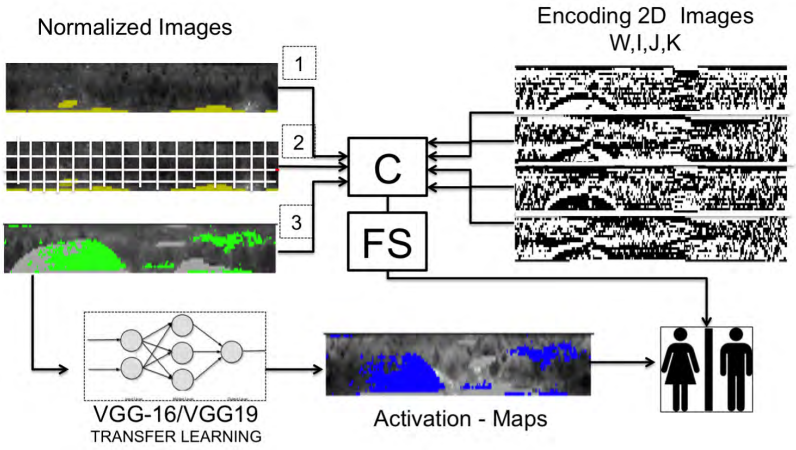

FIGURE 2. Block Diagram proposal. Number 1-represents the normalized image, 2-the Block features, and 3-the Pixel Features. C represents the Classifiers (SVM and 10 Ensemble). FS represents the Feature Selection Method. W, I, J, and K represent the real image, and encoding quaternionic images respectively.

specular reflections. All the pixels that belong to the mask have the same pixel value and therefore high redundancy information. Thus, the implementation that is used in this work generates an iris-normalized image that is $20 \times 240 = 4,800$ features. All features were used with an intensity value using 8-bit/pixel.

Traditionally, the iris 2D normalized pattern is broken up into a number of 1D signals, usually 20 rows for gender, and then these 1D signals are convolved with 1D Gabor filters with the following parameters: The wavelength (in pixels) of the log Gabor filter is 18 pixels and the ratio (σ/f_0) of the log Gabor filter is 0.5, where σ denotes bandwidth and f_0 denotes the central frequency, as in [26].

III. METHOD

This paper presents five experiments: Experiment 1: we used the whole features from normalized images to classify gender. This experiment was used as a baseline. Experiment 2: we used a transfer learning approach with a VGG19 model as feature extractor for gender classification. Experiment 3: we selected the most predominant blocks for gender classification from normalized images using a genetic algorithm (GA). Experiment 4: we selected the most predominant pixels using p-values, Conditional Mutual Information (CMI), Condition Mutual Information Maximization (CMIM) and Random Forest (RF) as feature selection methods. Experiment 5: we encoded the images using a Quaternionic-Code with 4 bits per pixels for classifying gender. Also, we selected the most relevant features from the four result images (one real and three imaginary) in order to reduce the number of features. In Figure 2, we shown a block diagram of the experiments.

A. ALL FEATURES

To classify gender, we used a normalized image from the left and right iris images separately. Each image has a size of $1 \times 4,800$ because we used only the real part of encoding image [18]. We transform each image into a vector of a matrix, M. Each row of M represents one normalized image and each column in M represents one feature. To analyze the different approaches, we used ten ensemble classifiers with the normalized image: AdaboostM1, LogitBoost, GentleBoost, RobustBoost, LPBoost, TotalBoost, RUSBoost with ''Tree learners'' classifiers and learning rate of 0.1, Random Forest classifier (RF) with 900 ''Trees'', and also an SVM classifier with Gaussian kernel with LIBSVM implementation [12].

B. TRANSFER LEARNING

We compared the hand-crafted feature extraction method with automatic Deep Learning-feature extraction. In the latter, we used transfer learning techniques since data is not available for direct training of the deep neural network. In many real-world applications, such as biometrics, data may not be available in the same domain for training, but is available in another domain, e.g., there is sufficient training data in Imaginet, the coco dataset [35], where the latter data may be in a different feature space or follow a different data distribution. In such cases, knowledge transfer, if done successfully, would greatly improve the performance of learning by avoiding many expensive data labeling efforts [34].

In our particular case, the VGG-19 architecture was trained on the Imagenet dataset which contains a total number of 1,000 different classes. Hence, these models are not expected to have learned normalized iris features that are relevant to the gender-classification problem. However, a model trained on a large dataset will contain many learned basic features, such as edges, spots, ridges, or horizontal lines that might be transferable to the iris datasets. Only the convolutional part of the model is instantiated, i.e., everything up to the fully-connected layers. Subsequently, the model is run on our training and testing data once, recording the output in

two arrays, i.e., the "bottleneck features" from the VGG-19 model, the last activation maps before the fully-connected layers. We, therefore, used the VGG-19 as a feature extractor. Then, we used the bottleneck features to initialize the fine-tuning approach.

C. BLOCK SELECTION

Block selection is based on GAs and operate iteratively on a population of structures, each one of which represents a candidate solution to the problem at hand, properly encoded as a string of symbols (binary). A randomly generated set of such strings forms the initial population from which the GA starts its search. Three basic genetic operators guide this search: Uniform selection, crossover in a single bit, and mutation. The genetic search process is iterative: evaluating, selecting, and recombining strings in the population during each iteration (generation) until reaching some termination condition.

Evaluation of each string is based on a fitness function that is problem-dependent. It determines which of the candidate solutions are better. Selection of a string, which represents a pixel in the search space, depends on the string's fitness relative to that of other strings in the population. The GA probabilistically removes from the population those blocks that have relatively low fitness.

Mutation, as in natural systems, is a very low probability operator and just flips a specific bit. Mutation plays the role of restoring lost genetic material. Crossover, in contrast, is applied with high probability. It is a randomized yet structured operator that allows information exchange between points. The goal is to preserve the fittest individuals without introducing any new value. In this paper we change the mutation rate from 0.1 up to 0.001 and cross-over rate was set up to 0.8.

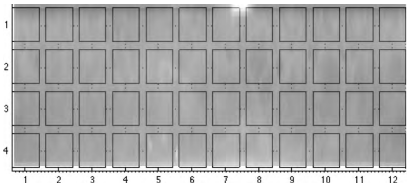

FIGURE 3. Normalized image divided in blocks of size 5 × 20. In total we have 48 blocks (4 rows and 12 columns).

1) CODIFICATION

Each image is processed starting from the normalized images of size 20 × 240 that is divided into blocks of 5 × 20 without overlap (This window size reached the best performance) and then concatenated in a vector of images. Therefore, we have 48 blocks for each image (4 rows × 12 columns). See Figure 3. In our encoding scheme, the chromosome is a bit string whose length is determined by the number of blocks or areas. Each block is associated with one bit in the string. If the *ith* bit is 1, then the *ith* block is selected, otherwise, that component is ignored. Each chromosome thus represents a different block subset.

2) INITIAL POPULATION

In general, it is common to generate the initial population randomly, (e.g., each bit in an individual is set randomly). In this way, however, we will end up with a population where each individual contains the same number of 1's and 0's on the average. To explore subsets of different numbers of features, the number of 1's for each individual is generated randomly. Then, the 1's are randomly scattered in the chromosome. In this work, we changed the population from 100 up to 500.

3) FITNESS EVALUATION

The goal of feature subset selection is to use fewer features to achieve the same or better performance. Therefore, the fitness evaluation contains two terms: (i) accuracy from the validation data and (ii) number of block used [36] . Only the features in the block subset encoded by an individual are used to test a classifier. The performance of the classifier is estimated using a validation data set and used to guide the GA. Each block subset contains a certain number of features. If two subsets achieve the same performance, while containing a different number of blocks, the subset with fewer blocks is preferred. Between accuracy and block subset size, accuracy is our major concern. Combining these two terms, the fitness function is given as:

$$fitness = 10^4 \times Accuracy + 2,08 \times Nb, \tag{1}$$

where accuracy is the classification rate achieved by an individual from an SVM previously trained, and Nb is the number of blocks selected. The accuracy ranges roughly from 0.5 to 1 (i.e., the first term assumes values in the interval 5,000 to 10,000). The number of block selected Nb ranges from 0 to L where L is the length of the chromosome. The second term takes values in the interval 0 to 100 since $L = 48$.

Overall, higher accuracy implies higher fitness. Also, fewer features used imply a greater number of zeros, and as a result, the fitness increases. It should be noted that individuals with higher accuracy will outweigh individuals with lower accuracy, no matter how many features they contain.

D. FEATURE SELECTION

Feature selection can be classified into three main groups: Filters, Wrappers, and Embedded [21]. Feature selection is closely related to feature extraction, a process in which feature vectors are created from the original dataset through manipulations of the data space, and can be considered to be a superset of the feature selection techniques.

Feature selection is also a broad field in continuous evolution, since the problem of selecting the most relevant and non-redundant features has not been solved for complex problems, such as in gender classification [39]. Eliminating relevant or non-redundant features would result in poor behavior of the classifier [39]. We used a strategy for selecting the most important features (relevant non-redundant) using univariate, bi-variate, and multi-variate methods based on statistical tests.

1) UNIVARIATE

The main idea of this stage is to complement and to reduce the dimension of the data by finding a small set of important features which can give good classification performance considering the previous block selection approach. Filters are usually used as a pre-processing stage since they are simple and fast. A widely - used filter method for data is to apply an univariate criterion separately on each feature, assuming that there is not interaction between features, we apply the statistical t-test approach on each feature and compare the p-value ($p < 0.05$) for each feature as a measure of how effective it is at separating features from female and male iris images. In order to get a general idea of how well-separated become the two gender groups (Female and Male), the analysis is performed by each feature. See Figure 4.

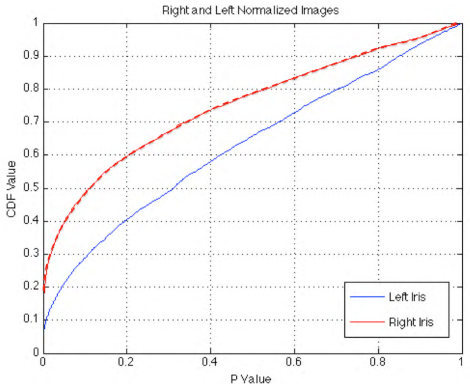

FIGURE 4. Empirical Cumulative Distribution function of the p-values for the right (dot-line) and left iris.

Figure 4, shows in blue (left iris) that there are about 10% of features having p-values close to zero and over 20% of features having p-values smaller than 0.05 meaning there are more than 960 features among the original 4, 800 features that have strong discriminant power.

Figure 4, also shows in red (right-iris, dot-line) that there are about 20% of features having p-values close to zero and over 40% of features having p-values smaller than 0.05 meaning there are more than 1,920 features among the original 4, 800 features that have strong discriminant power. We can sort these features according to their lower p-values and select some features from the sorted list.

2) BI-VARIATE

The previous selection algorithm did not consider the interaction between features, that means that features selected from a list based on their individual ranking of p-values may also contain redundant information, therefore not all the features are needed. This kind of simple feature selection approach is usually used as a pre-processing step since it is fast to apply. More advanced feature selection algorithm might improve the performance. Sequential feature selection selects a subset of features by sequentially adding (forward search) or removing (backward selection) until the stop condition criterion is reached. We use a forward sequential feature selection in a wrapper approach based on *MI* to find the most important features. Thus we use *MI* to measure the relationship between pairs of features.

MI is defined as a measure of how much information is jointly contained in two variables [15], or the degree to which knowledge of one variable determines the other variable. *MI* forms the basis of information-theoretic feature selection, as it provides a function for computing the relevance of a variable with respect to the target class (male or female). *MI* has two main properties. First, it can measure any kind of relationship among random variables, including nonlinear relationships. 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.

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

$$MI(x, y) = \int \int p(x, y) log \frac{p(x, y)}{p(x)p(y)} dxdy. \tag{2}$$

3) MULTIVARIATE-VARIATE

There is a limit to the level of performance that can be extracted from bi-variate feature selection methods [46]. This is because of feature complementarities, which means that more information can be found through the use of multiple features (three or more) than using features individually [48]. We used three methods with complementarity properties such as: Conditional Mutual Information Maximization (*CMIM*) [20], Weighted Conditional Mutual Information Maximization ($W - CMIM$) [39] and, Random Forest method (*RF*).

The *CMIM* criterion is a tri-variate measure of the information associated with a single feature about the class, conditioned upon an already selected feature [20]. It loops over the selected features and assigns each candidate feature a score based upon the lowest *CMI* between the selected features, the candidate feature, and the class [20], [21]. Then, the selected feature is the one with the maximum score.

$$CMIM = \begin{cases} arg\ max_{fi \in F} \{MI(fi; c)\} for\ S = \emptyset \\ arg\ max_{fi \in F/S} \{min_{fj \in S} MI(fi; c/fj)\} for\ S \neq \emptyset, \end{cases} \tag{3}$$

where, we compute the (*MI*) between the candidate variable *fi* and the output class *c* given each one of the variables in the set *S*, separately. This measure allows preserving a certain tradeoff between the power prediction of *fi* with respect to the output and the independence of the candidate feature with each one of the variables previously selected, considering relevance and redundancy.

Random Forest (RF) also can be used to select features. Random forest consists of a number of decision trees. Every node in the decision tree is a condition on a single feature, designed to split the dataset into two so that similar

response values end up in the same set. The measure based on which the (locally) optimal condition is called impurity. For classification, it is typically either Gini impurity (GDI), Two Deviance Criterion (TDC) or Twoing Rule (TR) and for regression trees it is variance [7]. Thus when training a tree, it can be computed how much each feature decreases the weighted impurity in a tree. For a forest, the impurity decrease from each feature can be averaged and the features are ranked according to this measure. RF also can be used when there are many more variables than observation, has a good predictive performance, incorporates interaction among predictor variables and returns measures of variable (pixels or feature) importance.

The Gini's Diversity Index (GDI):

$$1 - \sum_{i=1} = p^2(i), \tag{4}$$

where, the sum is over the classes i at the node, and $p(i)$ is the observed fraction of classes with class i that reach the node. A node with just one class (a pure node) has Gini index 0; otherwise, the Gini index is positive. Therefore, the Gini index is a measure of node impurity.

Also, we can use the TDC, with $p(i)$ defined the same as for the Gini index, the deviance of a node is:

$$-\sum_{i=1} = p(i)logp(i), \tag{5}$$

and the TR:

$$P(L)P(R)(\sum \mid L(i) - R(i) \mid)^2, \tag{6}$$

where $P(L)$ and $P(R)$ are the fractions of observations that split to the left and right respectively. If the expression is large, the split made each child node purer. Similarly, if the expression is small, the split made each child node more similar to each other, and hence similar to the parent node, and so the split does not increase node purity.

To select features (pixels) we can iteratively fit random forests, at each iteration building a new forest after discarding those variables (pixels) with the smallest variable significance; the selected set of pixels is the one that yields the smallest error rate. Random forest returns a measure of error rate based on the out-of-bag (OOB) cases for each fitted tree, the OOB error, and this is the measure of error we will use. Note that in this paper we are using OOB error to choose the final set of features (pixels), not to obtain unbiased estimates of the error rate of this rule.

E. QUATERNIONIC-CODE

In this paper, we report enhancing iris feature extraction by using a 2-D quadrature filter instead of the typical 1D Gabor filter. After filtering, a phase quantization step is used to generate the template images for gender classification. In this approach, phase information is encoded using the 2D Quaternioc Quadrature Filter (QQF) [11]. As will be shown, our experimental results indicate that the performance of the QQF applied to gender classification outperforms the traditional 1D Gabor filters. The Quaternionic-Code is essentially a binarized phase representation of an image using QQFs [5]. The use of quadrature filters is motivated by their ability to localize spatial and frequency domain phase information jointly in normalized iris images.

IV. DATASET

The images used in this paper were the same used in [39] and were taken with an LG 4000 sensor. The LG 4000 uses near-infrared illumination and acquires a 480x640, 8-bit/pixel images. Example of LG 4000 iris images appears in Figure 5. The image dataset for this work is person-disjoint and consists of one left eye image and one right eye image for each of 750 males and 750 females, for a total of 3,000 images. It is important to note that a person disjoint database assures that iris images from the same person are only present in the training or in the testing partition, but not in both. Of the 1,500 distinct persons in the dataset, visual inspection of the images indicates that about 1/4 are wearing clear contact lenses. See Figure 5.

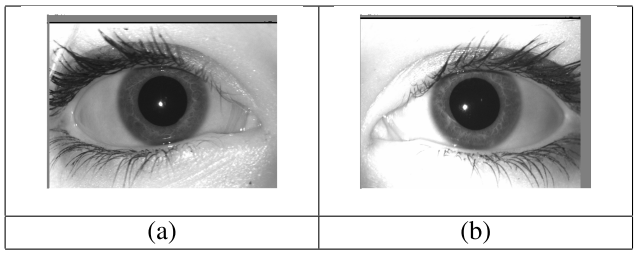

FIGURE 5. Examples of the LG 4000 iris images. Figure a) represents a left iris female image. Figure b) represents the right eye male image.

A training portion of the 1,500-person dataset was created by randomly selecting 60% of the males and 60% of the females. We used 5-fold cross-validation on this training set of 60% 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 60% training data, and a single evaluation is made on the 40% test 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.

A. VALIDATION DATASET

To validate our results we used the same validation set that [39]. The 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.

TABLE 2. Gender classification rates for the left and right iris image using normalized images for 10 different classifiers. Each result shows the standard deviation. The best results are highlighted in bold.

Method	Left (%)	Right (%)
SVM	62.00 +/- 0.25	63.50 +/- 0.30
Bag	**63.67 +/- 0.39**	**64.33 +/- 0.41**
AdaBoost M1	63.33 +/- 0.54	61.00 +/- 0.46
LogitBoost	60.00 +/- 0.5	62.33 +/- 0.34
GentleBoost	60.33 +/- 0.43	62.33 +/- 0.35
RobustBoost	57.33 +/- 0.65	56.00 +/- 0.54
LPBoost	54.67 +/- 0.41	62.67 +/- 0.43
TotalBoost	63.00 +/- 0.46	56.00 +/- 0.46
RusBoost	63.33 +/- 0.71	64.00 +/- 0.56
RF GDI	61.67 +/- 0.34	64.00 +/- 0.42

V. EXPERIMENTS AND RESULTS

A. USING ALL THE FEATURES - EXPERIMENT 1

Table 2 shows the results of gender classification when we used the whole normalized images ($1 \times 4, 800$). We used these results as a baseline to compare and measure the improvements of the feature selection methods. The first column of Table 2 shows the method used to classify the images. The second and third column shows the accuracy on gender classification rate for the left and right iris.

TABLE 3. Results of bottleneck and transfer learning approach with VGG-19. The best results are highlighted in bold.

Method	Dense	Bottleneck		Fine-tuning	
		Left (%)	Right (%)	Left (%)	Right (%)
Fine-Tunning	16.382	62.45	63.25	67.75	68.45
	8192	68.45	68.35	71.95	72.25
	4096	**72.25**	**74.55**	**77.45**	**79.65**
	2048	70.65	72.25	74.35	75.65
	1024	69.45	70.35	74.45	75.15
	512	68.15	69.45	72.35	73.45
	256	67.95	69.35	72.75	74.05
	128	67.35	68.95	71.35	73.65
	64	67.15	66.45	70.25	70.35
	32	61.35	65.35	64.45	69.15
	16	61.25	63.35	65.05	68.75

B. USING TRANSFER LEARNING - EXPERIMENT 2

Table 3 shows the results of gender classification when we used transfer learning approach with VGG-19. We used these results as a baseline to compare and measure the improvements of the feature selection methods. The first column of Table 3 shows the method used to classify the images. The second column shows the number of dense layers. The third and fourth columns show the accuracy on gender classification rate for the left and right iris using bottleneck extraction method. The fifth and sixth columns show the results when we used the fine-tuning approach.

C. USING BLOCKS SELECTION - EXPERIMENT 3

To classify gender using block selection we made a test with the GA, we used three different feature extraction methods,

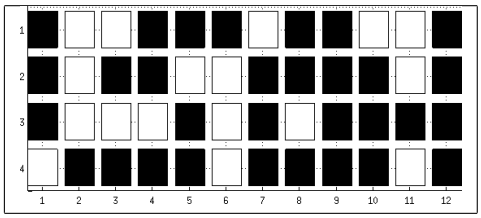

FIGURE 6. Example of the blocks selected from normalized image. The white blocks represent the most useful blocks selected. The black blocks represent the areas that are not relevant.

TABLE 4. Gender classification results, using Block selection with GA. The best results are highlighted in bold.

Method	Mut. Rate	String Size	Gen. Number	Accuraccy (%)
Right Eye				
Raw	**0.1**	19	50	**73.66**
Raw	0.01	19	50	70.66
Raw	0.001	19	50	71.66
PCA	**0.1**	20	50	**73.00**
PCA	0.01	20	50	71.00
PCA	0.001	20	50	70.00
LBP	**0.1**	17	50	**76.32**
LBP	0.01	17	50	73.32
LBP	0.001	17	50	75.32
Left Eye				
Raw	0.1	19	50	70.00
Raw	0.01	419	50	67.33
Raw	**0.001**	19	50	**72.00**
PCA	0.1	20	50	69.66
PCA	**0.01**	20	50	**71.66**
PCA	0.001	20	50	70.66
LBP	**0.1**	417	50	**76.67**
LBP	0.01	17	50	73.66
LBP	0.001	17	50	74.66

Raw pixels values, Principal Component Analysis (PCA) and Local Binary Patterns (LBP) to represent different features. In order to extract Raw and LBP features, we divide the images into 48 blocks. For PCA we used up to 50 eigenvalues vectors to rebuild the iris images. The best result was reached with LBP with 76.67% to rebuild the iris images. Table 4 shows the best results for RAW data, PCA and LBP. This table also shows the best parameters selected. To classify the images and compute the fitness we used an SVM classifier with RBF kernel. In Figure 6 we show the best blocks selected from the normalized image. The white squares represent the most useful blocks selected. The black squares represent the blocks that are not relevant. Some of the black blocks are localized in zones where the masked information is located. This mask is used at the initial stage for segmentation (see Figure 6).

D. FEATURE SELECTION METHOD - EXPERIMENT 4

1) USING A UNIVARIATE METHOD

We used the 1,000 most relevant features selected for the left iris and the 2,000 most relevant features for the right iris (In according with Figure 4). With the purpose of comparing the results with those of Table 2 (baseline) using again the ten ensemble classifiers and an SVM the accuracy of gender classification was re-estimated. We present the results in Table 5.

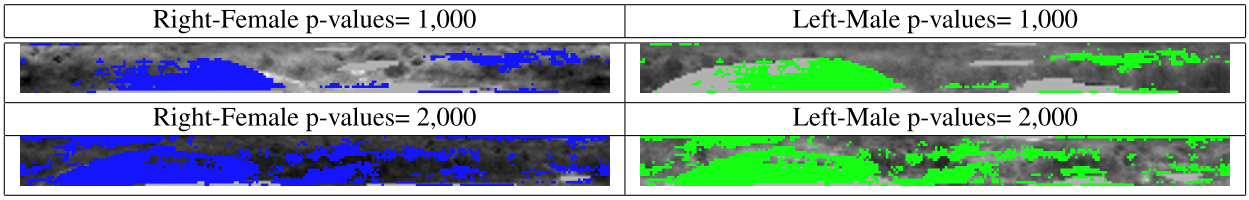

FIGURE 7. Features (pixels) over a particular normalized image with the lower 2,000 p-values for the right iris and the 1,000 lower p-values for left iris.

TABLE 5. Gender classification rates for the left and right iris image using the most relevant features in according to the p-values < 0.05 (Univariate method). The best results are highlighted in bold.

Method	Left (%)	Right (%)
SVM	61.35 +/- 0.37	62.00 +/- 0.34
Bag	**62.33 +/- 0.6**	**65.00 +/- 0.54**
AdaBoost M1	60.67 +/- 0.54	62.33 +/- 0.58
LogitBoost	60.67 +/- 0.37	61.33 +/- 0.41
GentleBoost	61.00 +/- 0.47	60.33 +/- 0.49
RobustBoost	60.67 +/- 0.33	63.67 +/- 0.30
LPBoost	59.00 +/- 0.65	59.67 +/- 0.56
TotalBoost	59.00 +/- 0.89	60.33 +/- 0.6
RusBoost	54.67 +/- 0.67	57.67 +/- 0.71
RF GDI	60.39 +/- 0.35	57.20 +/- 0.33

TABLE 6. Gender classification rates for the left and right iris image using the most relevant features in according to the *MI* (Bi-Variate method). The best results are highlighted in bold.

Method	Left (%)	Right (%)
SVM	65.53 +/- 0.37	64.00 +/- 0.36
Bag	64.30 +/- 0.6	67.00 +/- 0.53
AdaBoost M1	63.23 +/- 0.34	65.30 +/- 0.57
LogitBoost	62.67 +/- 0.29	63.30 +/- 0.41
GentleBoost	64.00 +/- 0.33	64.33 +/- 0.38
RobustBoost	64.65 +/- 0.31	65.67 +/- 0.27
LPBoost	62.00 +/- 0.65	63.67 +/- 0.26
TotalBoost	68.00 +/- 0.89	66.33 +/- 0.55
RusBoost	**68.67 +/- 0.67**	**68.67 +/- 0.47**
RF GDI	67.39 +/- 0.35	67.20 +/- 0.26

The localization of the selected features over a particular normalized image is shown in Figure 7. Clearly, the rank of features based on p-values did not improve the results. We show in gray color in Figure 7 the selected redundant features such as those in the masked areas.

2) USING A BI-VARIATE METHOD

We consider *MI* based for discrete (categorical) feature variables, the integral operation reduces to summation. In this case, computing *MI* is straightforward because both joint and marginal probability tables can be estimated by tallying the samples of categorical variables in the data. Finally, we used the 1,000 most relevant features selected for the left iris and the 2,000 most relevant features for the right iris, to re-estimated the accuracy of the gender classification, in order to compare the result with the Table 2 (baseline) using again the ten ensemble classifiers and SVM. We present the results in Table 6.

TABLE 7. Gender classification rates for the left and right iris image using the most relevant features selected by *CMIM* (Multivariate method). NFea represents the number of best features selected in step of 50. The best results are highlighted in bold.

Method	Left (%)	NFea	Right (%)	NFea
SVM	64.66 +/- 0.35	900	64.66 +/- 0.31	1,600
Bag	64.67 +/- 0.6	900	68.67 +/- 0.34	1,650
AdaBoost M1	63.33 +/- 0.54	800	63.67 +/- 0.42	1,300
LogitBoost	64.67 +/- 0.37	900	65.00 +/- 0.40	1,300
GentleBoost	64.00 +/- 0.47	600	65.33 +/- 0.39	1,400
RobustBoost	64.00 +/- 0.33	1000	68.33 +/- 0.30	1,700
LPBoost	65.33 +/- 0.65	700	66.33 +/- 0.46	650
TotalBoost	64.00 +/- 0.89	600	65.67 +/- 0.6	1,800
RusBoost	67.67 +/- 0.67	600	68.67 +/- 0.51	1,400
RF GDI	74.67 +/- 0.35	500	78.33 +/- 0.33	1,650
RF TDC	74.33 +/- 0.31	600	77.00 +/- 0.35	1,500
RF TR	**79.33 +/- 0.32**	700	**79.33 +/- 0.30**	1,900

3) USING A MULTI-VARIATE METHOD

The *CMIM* criterion selects relevant variables, avoids redundancy and, unlike previous methods, does not ignore variable complementarity. We made a test using the features (pixels) selected by *CMIM* in step of 50 until reach 4, 800, in order to compare the results with the Table 2 (baseline) using the ten ensemble classifiers and SVM again. We present the results in Table 7.

The $CMIM - W$ used in [39] considers that the feature fi is relevant only if it provides high information about the gender, considering the synergy and complementarity [46]. In this method, we compute the relevance and redundancy, adding the weight information, therefore, $wi(fi)$ is used directly instead of the feature fi. The weight information represents the relationships among one random image and its neighboring images. In this paper, we defined three parameters: t represents the number of times the process is repeated, p the number of images selected, and n the number of features selected from the matrix of all images. We changed the number of t in steps of 5 up to 10, and n in steps of 100 features up

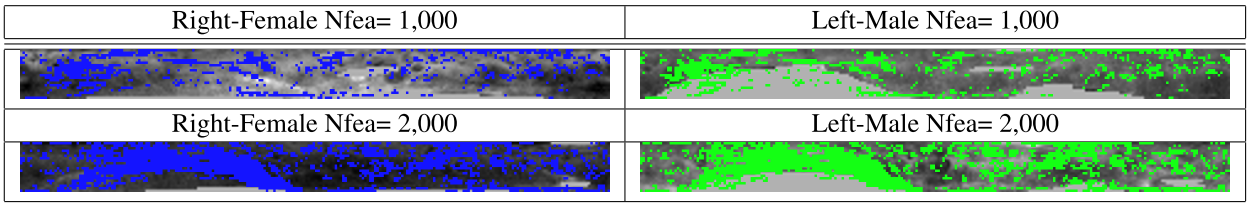

FIGURE 8. Features (pixels) over a particular normalized image with the best 2,000 features selected (Nfea) by Random Forest for the right iris and the best 1,000 features (NFea) for left iris.

TABLE 8. Gender classification rates for the left and right iris image using the most relevant features selected by *CMIM* − *W* (Multivariate method). NFea represent the number of best features selected in step of 50 up to 1,000 for the left and 2,000 for the right. The best results are highlighted in bold.

Method	Left (%)	NFea	Right (%)	NFea
SVM	69.00 +/- 0.35	600	69.33 +/- 0.31	1,000
Bag	69.33 +/- 0.6	600	72.37 +/- 0.34	1,050
AdaBoost M1	68.66 +/- 0.54	700	68.37 +/- 0.42	1,100
LogitBoost	69.33 +/- 0.37	600	70.00 +/- 0.40	1,100
GentleBoost	69.33 +/- 0.47	700	70.66 +/- 0.39	1,000
RobustBoost	69.66 +/- 0.33	700	73.00 +/- 0.30	1,300
LPBoost	70.00 +/- 0.65	500	71.00 +/- 0.46	650
TotalBoost	69.33 +/- 0.89	500	70.33 +/- 0.6	1,200
RusBoost	74.66 +/- 0.67	500	73.33 +/- 0.51	1,100
RF GDI	79.00 +/- 0.35	500	83.00 +/- 0.33	1,250
RF TDC	79.66 +/- 0.31	500	83.66 +/- 0.35	1,100
RF TR	**84.66 +/- 0.32**	700	**85.66 +/- 0.30**	1,000

TABLE 9. Gender classification rates for the left and right iris image using the most relevant features according to the RF Ranking (Multivariate method). NFea represents the number of best features selected in step of 50. The best results are highlighted in bold.

Method	Left (%)	NFea	Right (%)	NFea
SVM	65.66 +/- 0.35	900	66.67 +/- 0.35	1,100
Bag	64.33 +/- 0.6	800	68.33 +/- 0.54	1,550
AdaBoost M1	66.00 +/- 0.54	550	66.67 +/- 0.58	1,700
LogitBoost	63.33 +/- 0.37	500	66.67 +/- 0.41	1,300
GentleBoost	63.33 +/- 0.47	800	68.67 +/- 0.49	1,950
RobustBoost	64.33 +/- 0.33	650	68.67 +/- 0.30	1,700
LPBoost	64.33 +/- 0.65	550	66.00 +/- 0.56	1,300
TotalBoost	62.67 +/- 0.89	500	64.67 +/- 0.6	1,850
RusBoost	64.67 +/- 0.67	750	66.00 +/- 0.71	1,050
RF GDI	74.00 +/- 0.35	400	78.67 +/- 0.33	1,300
RF TDC	75.00 +/- 0.31	850	79.33 +/- 0.31	1,500
RF TR	**76.33 +/- 0.32**	800	**78.33 +/- 0.30**	1,900

to 4,800. We also explored values of p from 5 up to 50 image classes, searching for the value that would produce the best classification rate using a forward and backward selection.

The $CMIM-W$ criterion selects relevant groups of images, avoids redundancy and, unlike previous methods, uses the complementarity information. To compare the results those of Table 2 (baseline), we used the ten ensemble classifiers and SVM again. We present these results in Table 8. This approach reaches the best classification rate instead of p-value, *MI*, *RF*, and *CMIM* approaches.

RF was used for the purpose of comparing our results to those shown in Table 2 to 8. The ten ensemble classifiers and SVM again were used, but we selected features in steps of 50 up to 4,800. We present the results in Table 9. The best result in Table 9 was reached by RF using the TR split criterion.

In Figure 8, the distribution of the best features selected by RF over a particular image in the background was used. This approach reduces the relevance information using less number of features, but also with high redundancy selecting masked area. This method reaches similar results as those reached by *MI* but with less number of features selected.

In Figure 9, we show the distribution of the best features selected by $CMIM - W$ over a particular image in the background. This method using complementary information that allows to reduce the trade-off between relevance and redundancy selected the lower number of features with higher accuracy. The difference with Figure 7 and Figure 8 is evident. The images also show that the best features did not consider the mask area.

To validate the results we use the validation set described in section 4.A. We evaluated the best method of Table 8 with these images because this method reached the best results for each experiment in according with the Table 2, 3, 4, 5, 6 and 7. The results are presented in Table 10.

E. USING A QUATERNIONIC-CODE - EXPERIMENT 5

In order to improve our results, we classified gender using the encoding images with Quaternioc-Code(QC) with 3 and 4 bits per pixel. The best result was reached with 4 bits per pixel. The QC produces four images as results; one image for the real part, (i), and three images, (j, k, l), for the imaginary parts. See Figure 10.

The results with only real part encoding images are presented in Table 11. The best result was reached by GentleBoost with 79.45% and 77.45% for the left and right irises. These results outperform the best results of Table 2 by up to 16% using the normalized images. However, these results are not better than those reported in Table 8 with $CMIM - W$ and Random Forest. The best results

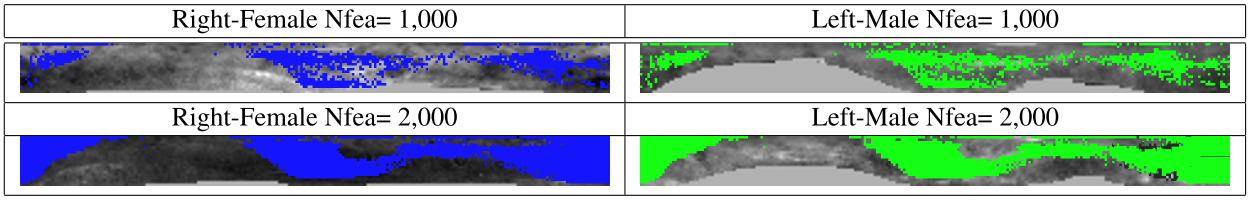

FIGURE 9. Features (pixels) over a particular normalized image with the best 2,000 features selected (Nfea) by $CMIM - W$ for the right iris and the best 1,000 features (Nfea) for left iris.

TABLE 10. Gender classification rates for the left and right iris image using the most relevant features selected by $CMIM - W$. NFea represents the number of best features selected in step of 50. The best results are highlighted in bold.

Method	Left (%)	Val	Right (%)	Val
SVM	69.00 +/- 0.35	66.00	69.33 +/- 0.31	67.33
Bag	69.33 +/- 0.6	65.66	72.37 +/- 0.34	70.33
AdaBoost M1	68.66 +/- 0.54	65.33	68.37 +/- 0.42	66.33
LogitBoost	69.33 +/- 0.37	67.33	70.00 +/- 0.40	67.66
GentleBoost	69.33 +/- 0.47	66.66	70.66 +/- 0.39	68.66
RobustBoost	69.66 +/- 0.33	67.33	73.00 +/- 0.30	71.33
LPBoost	70.00 +/- 0.65	68.66	71.00 +/- 0.46	69.66
TotalBoost	69.33 +/- 0.89	67.33	70.33 +/- 0.6	68.33
RusBoost	74.66 +/- 0.67	71.33	73.33 +/- 0.51	71.66
RF GDI	79.00 +/- 0.35	76,66	83.00 +/- 0.33	73.66
RF TDC	79.66 +/- 0.31	77.33	83.66 +/- 0.35	81.66
RF TR	**84.66 +/- 0.32**	81.66	**85.66 +/- 0.30**	82.33

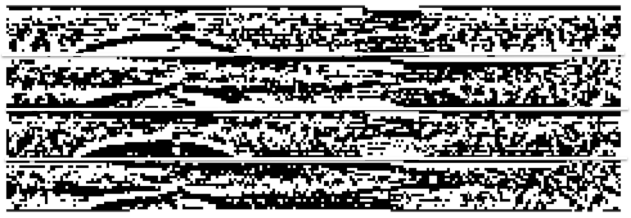

FIGURE 10. Quaternionic image results. Top row image represents the real part. Second, third and fourth rows represent the imaginary part.

overall were reached by $CMIM - W$ with 10 nearest neighbor images, (p), selected from the four images from the Quaternionic-Code. We explored the best number of features from 100 up to 4,800. A GentleBoost classifier reached 93.45% for the left and 95.45% for the right iris with 2,400 groups of features. Our results in Table 12 outperform by up to 25% compared to the baseline presented in Table 2, and almost 6% relative to the best results reported as the state of the art. The results in Table 12 also exceeded the best results in Table 2, Table 4, Table 5, Table 6, Table 7, Table 8, Table 9, Table 10 and Table 11.

To validate the results with encoding images, we used the validation set described in Section 4.A. We evaluated the best method of Table 12 with these images because when we compared the results of all the tables, this method achieved

TABLE 11. Quaternionic-Code results using only the real part encoding images. The best results are highlighted in bold.

Method	Left (%)	Right (%)
SVM	60.51	59.85
Bag	68.35	77.25
AdaBoost M1	70.35	77.05
LogitBoost	79.15	77.45
GentleBoost	**79.45**	**77.45**
RobustBoost	70.55	73.70
LPBoost	73.70	72.15
TotalBoost	73.45	73.95
RusBoost	79.45	78.65
RF GDI	77.30	79.05
RF GINI	74.85	73.35

TABLE 12. Quaternionic-Code results plus feature selection method using 4 bits encoding images. Number of neighbors p = 10. In the parenthesis we show the best numbers of feature selected with $CMIM - W$. The best results are highlighted in bold.

Method	Left (%)	Val (%)	Right (%)	Val (%)
SVM	84.65 (2,050)	81.35	87.35 (2,100)	85.15
Bag	83.45 (2,100)	80.55	81.25 (2,100)	80.15
AdaBoost M1	86.35 (2,100)	83.25	87.15 (2,100)	84.75
LogitBoost	89.15 (2,100)	86.55	87.45 (2,050)	85.65
GentleBoost	**93.45 (2,400)**	**91.15**	**95.45 (2,400)**	**93.15**
RobustBoost	90.55 (2,100)	88.65	93.55 (2,500)	91.75
LPBoost	93.60 (2,050)	91.15	92.45 (2,500)	89.65
TotalBoost	89.45 (2,100)	86.35	88.95 (2,600)	87.45
RusBoost	89.55 (2,100)	85.95	88.65 (2,000)	86.25
RF GDI	91.30 (2,300)	89,15	90.05 (2,500)	88.65
RF GINI	91.80 (2,300)	89.65	93.25 (2,700)	90.15

the best results for each experiment. The results are presented in Table 12.

Figure 11, show the most relevant features from the real and imaginaries images results from QQC. The features were selected separated in steps of 100 up to 4,800 from each image and then we merged the best 600 for each one of them. The results are presented in Table 12.

FIGURE 11. The most relevant features selected from each of the four QQC encoding images. Top row image represents the real part. Second, third and fourth rows represent the imaginary part. White pixels represent features selected.

TABLE 13. Show the time spent for each feature selection method in average to select the best 1,000 features in the standalone process using a python implementation. RF represents Random Forest. QQC represents Quaternionic-Code.

Feature Selection Method	Time (Sec.)
RF Ranking	300
CMIM	500
CMIM-W	640
QQC	430

VI. TIME COMPLEXITY

Table 13, shows the results in seconds spent by the four feature selection methods used in this research. The feature selection process is performed offline only one time, and the selected features are used to classify gender online. The reported time is an average of all of the extracted features. The time was computed using an Intel I7-6500U of the 6th generation 2.7 GHz, with 32 GB RAM and an Ubuntu 16.04 operative system. The feature selection methods were implemented in Python. A C++ implementation in a parallel processing architecture should reduce the computational time significantly.

The computational cost of the CMIM-W algorithm depends on the estimation of MI. The estimation of the MI is $O(N \times logN)$. Therefore, the time complexity is $O(2 \times M \times N \times logN)$, where 2 is the number of classes of the vector (Male, Female); M is the number of features in the set; and N is the number of image samples available.

VII. CONCLUSION

The methods and the proposed experiments in this research have the objective of finding the best features on the iris for classifying gender using NIR images. We explored the information from the entire image as a baseline to study the pixel distribution (Experiment 1). Then, an automatic feature extraction method based on transfer learning was proposed (Experiment 2). Later, using a GA, a block localization from the images was used to find the most relevant blocks of pixels that improve gender classification (Experiment 3). Afterwards, a feature selection method was proposed to analyze the quantity and the quality of information from univariate, bivariate, and multivariate feature selection methods (Experiment 4). Also, it was concluded, from state-of-the-art literature, that traditional methods lose much information when encoding the iris because these methods remove the imaginary part and lose part of the 2D information after applying 1D Gabor filters. The objective, therefore, was to design a method that could capture the spatial relationship in both directions in order to codify the iris information in a better way. QQC reached the best results, encoding the iris information in 4 bits per pixel (Experiment 5). This codification, along with the W-CMIM, allowed reaching the best gender classification rate.

Finally, the methods show that using all the pixels from the normalized iris images yields only suboptimal results. In addition, the traditional approach loses 2D information impeding achieving the best results. The 2D spatial information from the iris is relevant in improving gender classification.

According to our results, the proposed method is able to encode the phase information and the 2D spatial information from the normalized iris images using 4 bits per pixel with a 2D Gabor filter. This information is not available with the traditional approach using 1D Gabor filters. Additionally, the feature selection method improves the results of gender classification significantly, since using all the bits on the iris tends to confuse the classifier, as has been stated previously in the literature [3], [39], [46]. Feature selection enables eliminating noisy inputs and redundant features into the classifier. The proposed encoding has the ability to localize spatial and frequency domain phase information jointly on the normalized iris images. As a result of encoding the images with the QQC method, four images were obtained. Using only the real part was not enough to reach the best results, as is shown in Table 11. A group feature selection method was necessary to select the most relevant features from the four images, as is shown in Table 12.

In order to show that the number and position of features are representative and independent of the classifier, we tested our approaches using 10 popular machine learning classifiers, and also in a pre-trained model of a convolutional neural network using VGG-19 as the feature extractor.

The best result achieved in this work was with a Quaternionic-Code using 4 bits to codify each pixel. These results are the best for gender classification with the UND datasets and show that the Quaternionic-Code with the 2D quadrature filter approach outperforms previous results using a 1D approach.

Also, the results reported in this paper show the location of the most relevant blocks and the number of features available from the iris that improve the gender classification rate. The paper shows that gender information in the iris is not localized in a specific block. The results show that the relevant information is distributed across different areas, because the information of the iris is represented by minimum structures, i.e., groups of the most discriminative features, instead of isolated features. This information may shine some light on how biological information is present and organized on the iris.

The GA allowed us to find the minimum block size, 5×20, for classifying gender. The results also show that when a selection of the most relevant blocks or features from both the left and right irises was used, the classification rate was significantly better than when using all the information available from both irises.

The best results for gender classification were reached using complementary methods with normalized images instead of univariate or bivariate methods because the univariate and bivariate methods did not remove the redundancy of the data.

For the complementary methods using an RF classifier, $CMIM - W$, achieved the best results of 84.66% with 700 features, and 85.66% with 1,000 features for the left and right irises. This represents 15% of the number of features of the left iris and 21% for the right iris from the total of 4,800 pixels available. This approach also solved the problem of make-up present in the eye image. The best results used only the most relevant pixels and the features with make-up had lower accuracy and higher redundancy as is shown in Figure 7 and Figure 8.

For the 2D gabor filters, Quaternionic-Code, and $CMIM - W$ the best results reached were 92.45% with 2,400 features for the left iris, and 95.45% with 2,400 features for the right iris. This represents 30% of the number of features from the total of 4,800 bits.

This paper can be used as a guide to develop new methods to select features and improve the new soft biometrics problems such as: gender, age, and emotion from iris.

REFERENCES

[1] O. A. Alim and M. Sharkas, "Texture classification of the human iris using artificial neural Networks," in *Proc. 11th IEEE Medit. Electrotechnical Conf.*, May 2002, pp. 580–583.

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

[3] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35. no. 8, pp. 1789–1828, Aug. 2012.

[4] D. Bobeldyk and A. Ross, "Iris or periocular? exploring sex prediction from near infrared ocular images," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2016, pp. 1–7.

[5] D. Boukerroui, J. A. Noble, and M. Brady, "On the choice of band-pass quadrature filters," *J. Math. Imag. Vis.*, vol. 21, nos. 1–2, pp. 53–80, Jul. 2004.

[6] 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, May 2008.

[7] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[8] C. Busch, J. Tapia, and C. Rathgeb, "Sex-prediction from periocular images across multiple sensors and spectra, in press," in *Proc. 14th Int. Conf. Signal Image Technol. Internet Based Syst.*, Nov. 2018, pp. 241–258.

[9] Canpass. (1996). *Canadian Border Services Agency*. Accessed: Mar. 5, 2019. [Online]. Available: https://www.cbsa-asfc.gc.ca/prog/canpass/menu-eng.html

[10] M. Castrillón-Santana, J. Lorenzo-Navarro, and E. Ramón-Balmaseda, "On using periocular biometric for gender classification in the wild," *Pattern Recognit. Lett.*, vol. 2, pp. 181–189, Oct. 2016.

[11] S. Chan and A. Kumar, "Reliable ear identification using 2-D quadrature filters," *Pattern Recognit. Lett.*, vol. 33, pp. 1870–1881, Oct. 2012.

[12] 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.

[13] C. Chen and A. Ross, "Evaluation of gender classification methods on thermal and near-infrared face images," in *Proc. Int. Joint Conf. Biometrics (IJCB)*, Oct. 2011, pp. 1–8.

[14] M. D. Costa-Abreu, M. Fairhurst, and M. Erbilek, "Exploring gender prediction from iris biometrics," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2015, pp. 1–11.

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

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

[17] J. Daugman, "Iris recognition at airports and border-crossings," in *Encyclopedia Biometrics*. Boston, MA, USA: Springer, 2009, pp. 819–825. doi: 10.1007/978-0-387-73003-5.

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

[19] J. S. Doyle and K. W. Bowyer, "Robust detection of textured contact lenses in iris recognition using BSIF," *IEEE Access*, vol. 3, pp. 1672–1683, 2015.

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

[21] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, *Feature Extraction, Foundations and Applications, Studies in Fuzziness and Soft Computing*. New York, NY, USA: Springer, 2006.

[22] H. Han, C. Otto, X. Liu, and A. Jain, "Demographic estimation from face images: Human vs. machine performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 6, pp. 1148–1161, Jun. 2016.

[23] P. Jafari and F. Azuaje, "An assessment of recently published gene expression data analyses: Reporting experimental design and statistical factors," *BMC Med. Inform. Decision Making*, vol. 6, no. 1, p. 27, Jun. 2006.

[24] A. Kuehlkamp, B. Becker, and K. Bowyer, "Gender-from-iris or gender-from-mascara?" in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2017, pp. 1151–1159.

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

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

[27] J. Merkow, B. Jou, and M. Savvides, "An exploration of gender identification using only the periocular region," in *Proc. 4th IEEE Int. Conf. BTAS*, Sep. 2010, pp. 1–5.

[28] C. Perez, J. Tapia, P. Estevez, and C. Held, "Gender classification from face images using mutual information and feature fusion," *Int. J. Optomechatronics*, vol. 6, no. 1, pp. 92–119, 2012.

[29] P. J. Phillips *et al.*, "FRVT 2006 and ICE 2006 large-scale experimental results," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 831–846, May 2010.

[30] N. Popescu-Bodorin, V. E. Balas, and I. M. Motoc, "Iris codes classification using discriminant and witness directions," in *Proc. 5th Int. Symp. Comput. Intell. Intell. Inform. (ISCIII)*, pp. 143–148, Sep. 2011.

[31] H. Proença and J. C. Neves, "Deep-prwis: Periocular recognition without the iris and sclera using deep learning frameworks," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 4, pp. 888–896, Apr. 2018.

[32] X. Qiu, Z. Sun, and T. Tan, "Global texture analysis of iris images for ethnic classification," in *Proc. Int. Conf. Adv. Biometrics, ICB*, 2006, pp. 411–418.

[33] A. Rattani, N. Reddy, and R. Derakhshani, "Gender prediction from mobile ocular images: A feasibility study," in *Proc. IEEE Int. Symp. Technol. Homeland Security (HST)*, Apr. 2017, pp. 1–6.

[34] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. Lecun. (2018). "Overfeat: Integrated recognition, localization and detection using convolutional Networks" [Online]. Available: http://arxiv.org/abs/1312.6229

[35] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556

[36] Z. Sun, G. Bebis, X. Yuan, and S. J. Louis, "Genetic feature subset selection for gender classification: A comparison study," in *Proc. 6th IEEE Workshop Appl. Comput. Vis.*, Dec. 2002, pp. 165–170.

[37] N. Othman, B. Dorizzi, and S. Garcia-Salicetti, "OSIRIS: An open source iris recognition software," *Pattern Recognit. Lett.*, vol. 82, pp. 124–131, Oct. 2016. doi: 10.1016/j.patrec.2015.09.002.
[38] 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.
[39] J. Tapia, C. Perez, and K. Bowyer, "Gender classification from the same iris code used for recognition," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 8, pp. 1760–1770, Aug. 2016.
[40] J. E. Tapia, *Iris and Periocular Biometric Recognition*. London, U.K.: IET, 2017.
[41] J. E. Tapia and C. Aravena, *Deep Learner for Biometrics*. New York, NY, USA: Springer, 2017.
[42] J. E. Tapia, C. A. Perez, and K. W. Bowyer, "Gender classification from iris images using fusion of uniform local binary patterns," *Proc. Eur. Conf. Comput. Vision-ECCV, Soft Biometrics Workshop*, 2014, pp. 751–763.
[43] J. E. Tapia and I. Viedma, "Gender classification from multispectral periocular images," in *Proc. Int. Joint Conf. Biometrics (IJCB)*, Oct. 2017, pp. 1–9.
[44] V. Thomas, N. Chawla, K. Bowyer, and P. Flynn, "Learning to predict gender from iris images," in *Proc. 1st IEEE Int. Conf. Biometrics Theory, Appl., Syst.*, Sep. 2007, pp. 1–5.
[45] UIDAI. (2014). *Unique Identification Authority of India*. Accessed: Mar. 5, 2019. [Online]. Available: https://uidai.gov.in/about-uidai.html
[46] 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.
[47] D. Wang, H. Zhang, R. Liu, W. Lv, and D. Wang, "t-test feature selection approach based on term frequency for text categorization," *Pattern Recognit. Lett.*, vol. 45, pp. 1–10, Aug. 2014.
[48] P. L. Williams and R. D. Beer. (2010). "Nonnegative decomposition of multivariate information." [Online]. Available: https://arxiv.org/abs/1004.2515

JUAN E. TAPIA received the P.E. degree in electronics engineering from Universidad Mayor, in 2004, the M.Sc. 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. Also, he spent one year of internship at the University of Notre Dame. From 2016 to 2017, he was an Assistant Professor with Universidad Andres Bello. He is currently an R&D Director of electronic and electricity area at the Universidad Tecnologica de Chile-INACAP. His main research interests include pattern recognition and machine learning applied to softbiometrics, gender classification, feature fusion, and feature selection.

CLAUDIO A. PEREZ (M'90–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 the Ph.D. degree, in 1991. He was a Fulbright Student with The Ohio State University, where he obtained a Presidential Fellowship, in 1990. He was a Visiting Scholar with the University of California at Berkeley, in 2002, through the Alumni Initiatives Award Program from the Fulbright Foundation. He was the Department Chairman, from 2003 to 2006, and the Director of the Office of Academic and Research Affairs, School of Engineering, Universidad de Chile, from 2014 to 2018, where he is currently a Professor with the Department of Electrical Engineering. His research interests include biometrics, image processing applications, and pattern recognition. He is a Senior Member of the IEEE Systems, Man and Cybernetics Society and the IEEE-CIS Society.