ATTRIBUTE PREDICTION FROM NEAR INFRARED IRIS AND OCULAR IMAGES

By

Denton Bobeldyk

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Computer Science – Doctor of Philosophy

2019

ABSTRACT

ATTRIBUTE PREDICTION FROM NEAR INFRARED IRIS AND OCULAR IMAGES

By

Denton Bobeldyk

The iris is the colored portion of the eye surrounding the pupil. Images captured in the visible spectrum make it difficult for the rich texture of brown irides to be discerned; therefore, iris recognition systems typically capture an image in the Near Infrared (NIR) spectrum. The region surrounding the iris, the ocular region, is also captured by the sensor during the imaging process.

The focus of this thesis is on developing methods for predicting soft biometric attributes of an individual based on the iris and ocular components of the eye. In addition to attribute prediction, the effect of covariates on attribute prediction are also studied. Attributes considered in this work include gender, race and eye color. For the gender and race attributes, both the iris and surrounding ocular region are analyzed to determine which region provides the greatest gender cues. A regional analysis reveals that the iris-excluded ocular region provides a greater gender prediction accuracy than the iris-only region. This finding is of great significance as, typically, the iris-excluded ocular region is discarded by the iris recognition system. This research reinforces the need to retain the iris-excluded ocular region for additional processing. For race, it is shown that the iris-only region provides better prediction accuracy. In order to study the stability of the gender and race features, the impact of image blur on attribute prediction was also examined. It is observed that as the level of image blur increases, the race prediction accuracy decays at a much faster rate than that of gender. For eye color, the textual cues presented on the iris stroma are exploited to generate a discriminatory feature vector that is capable of distinguishing between two categories of eye color. The impact of image resolution on attribute prediction was also determined. A convolutional neural network architecture is presented that is capable of attribute prediction using images as small as 5×6, a mere 30 pixels. Experimental results suggest the possibility of deducing soft biometric attributes from low resolution images, thereby underscoring the feasibility of extracting these attributes from

poor quality images. Finally, the thesis explores the possibility of harnessing the feature vector used to predict one attribute (e.g., gender) in order to predict a different attribute (e.g. race). The ensuing experiments convey the viability of cross attribute prediction in the context of NIR ocular images.

In summary, this thesis provides insight into attribute prediction from NIR ocular images by conducting an extensive set of experiments.

ACKNOWLEDGEMENTS

Thanks to my sister Tammy for always being a great friend. I'd also like to thank my brother-in-law Shawn and their four kids Luke, Connor, Anna and Emma for being such supportive family members.

Thank you Dr. Eric Torng for being an excellent teacher and believing in me ($log_3(3^6) = ?$).

Thank you Dr. Xiaoming Liu for the excellent course on computer vision and providing such a comprehensive and accurate view of the computer vision field. Thank you also for allowing me to continue this education and taking me on as an independent study.

A very special thanks to the members of the iProbe Research Lab. Thank you for your fellowship and acceptance of me as one of your own. Despite the fact that we began our journey at different stages of our lives, you welcomed me in and we were able to share the ups and downs of this journey together. Thank you for the critiques, the company, the comiseration, the laughs and friendships that I will never forget. I feel truly blessed to have known each of you and I am a better person for it.

Thank you to Davenport University and the support they have given me during my PhD pursuit.

Special thanks to the iris researchers that have gone before me. I'm grateful for the opportunities I've had to meet with some of you at conferences and am grateful for your fellowship. I'm also grateful for the papers and books that you have produced that assisted me in my progression and my eventual small contribution to the field.

Thanks to my crew at the Wharf that provided a great distraction and fun times with friends.

The process of acquiring my PhD over 6 years deeply changed me as a person. I am very thankful to Michigan State University and their rigorous Computer Science PhD program that cultivated this transformation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Portions of the material in this chapter have been previously published in the journal IEEE Access

1.1 Object Attributes and Recognition

Automated object recognition from digital images has been a popular research problem in the computer vision community [2, 12]. Object recognition is the process of identifying objects in images through the use of an object classifier (see Figure 1.1). Recently, deep learning methods have been used for effective object classification as can be seen in [42, 71, 78, 79, 77]. An object classifier based on deep learning is typically trained with a large number of images, some of which have the target object in them (positive examples) while others do not have the target object (negative examples) (see Figure 1.2). Once such a classifier is trained, it can receive an input image and output a class label, thereby identifying the target object in the image (see Figure 1.3).

The use of such an automated object classifier - based on deep-learning or any other method -

(a) An image that contains a baseball   (b) An image that does not contain a baseball

Figure 1.1: Two sample images, one that contains the target object and one that does not. The image in (a) is referred to as a positive example, while the one in (b) is referred to as a negative example.
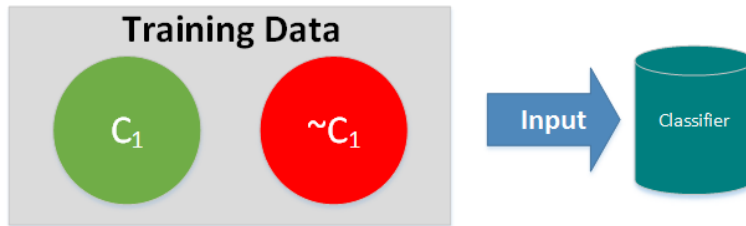
Figure 1.2: An object classifier is trained using positive and negative examples. The green circle represents positive images (contains the object class) while the red circle represents negative images (does not contain the object class).

Figure 1.3: A trained object classifier that receives test images as input and outputs a predicted object class label for each image.
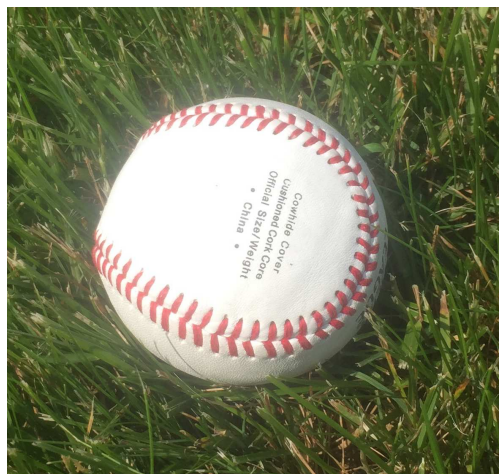
Figure 1.4: An image containing an object with a class label of 'baseball'. Sample attributes could be: white, round, pattern of red marks, red stitching, etc.

has several drawbacks: 1) the trained classifier will be unable to identify new classes of objects that were not previously seen (see Figure 1.5); 2) each time a new object class is added to the classifier, an additional training phase will be required. In order for the training to occur, a large number of positive and negative examples have to be selected and annotated. Annotation often entails marking the spatial extent of the target object in an image and indicating its label, i.e., the pre-defined name of the object. The labeling process can be time consuming and expensive. Lampert et. al [47] pose this problem as one with disjoint training/testing classes. The solution they present is to create an intermediate module that will predict *attributes* from the training images. An attribute is defined by [25] as 'visual qualities of objects, such as 'red', 'striped', or 'spotted'. They go on to state that attributes 'can be any combination of appearance, shape, or the layout of segments within the pattern'. Li et al.[49] provide further clarification by stating an attribute to be 'a high level semantically meaningful representation.' Figure 1.4 displays the image of an object (a baseball) with a list of sample attributes. Insertion of an intermediate module that is capable of predicting attributes could then be used to predict object classes based on the attributes that define that object (see Figure 1.6). Multi-task [1] and other variants of the attribute based model will not be discussed in this introductory section.

The attribute layer is composed of many different attribute classifiers. The attribute classifiers may each be trained separately as shown in Figure 1.7. Once the attributes are learned, the system will be capable of predicting objects based on attributes associated with the object (see Figure 1.6). Designing a system based on attribute classifiers simplifies the addition of a new class. A new class can be added by simply associating it with a list of pre-defined attributes.

Farhadi et al. [25] identify this solution as shifting the 'goal from naming to describing'. They also place an emphasis on 'discriminative' attributes. Semantic attributes like 'round' or 'white' may apply to both golf balls and baseballs, so a discriminatory attribute must be defined that separates the two classes of objects.

There are several advantages to utilizing an attribute based model of object classification summarized below [49, 25, 47]:
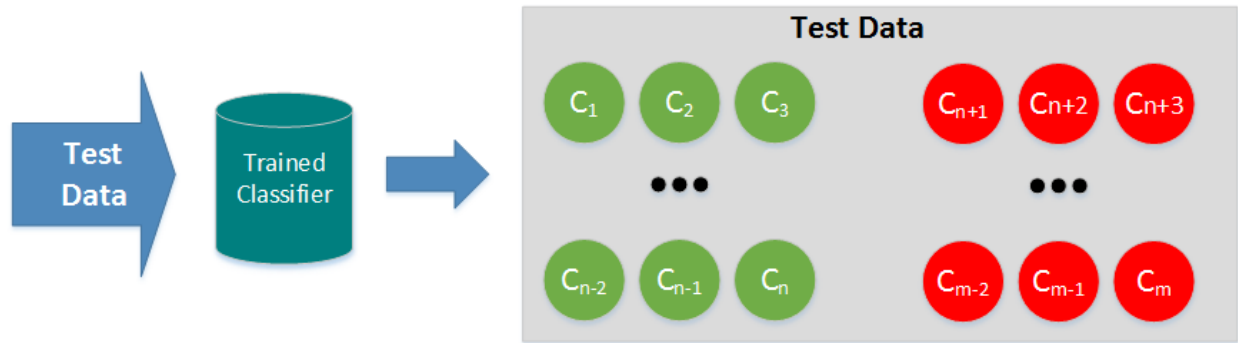
Figure 1.5: A trained classifier is unable to predict classes that were *not included* in the training data. The red circles represent images from classes that were not in the training data, the green circles represent images from classes that were included in the training data.
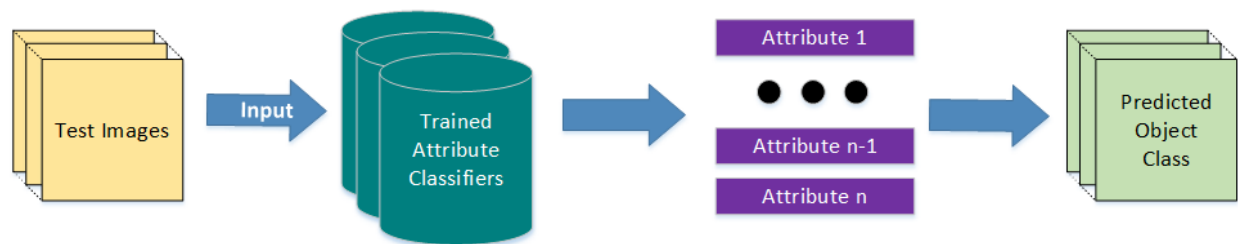
Figure 1.6: Test images are input to multiple attribute classifiers that generate a list of attributes. Each object class is defined by a set of attributes. Based on the set of attributes detected a class label is generated.

- Objects can be described by a semantic description, not simply given a label [49]

- New object classes can be learned simply by associating them with a set of attributes [25]

- Fewer total classifiers that require training are needed for an attribute based model. This is possible because each of the objects classes will share from a smaller pool of trained attribute classifiers [25].

- Large image collections for *each* object classifier are not needed for training. Image collections need only be used for the smaller number of attributes [47]

1.2   Person recognition and attribute prediction

People, like objects, can be described by their attributes. For example, a person could be described as a 'short brown-eyed male with long hair'. However, unlike objects, attributes may not
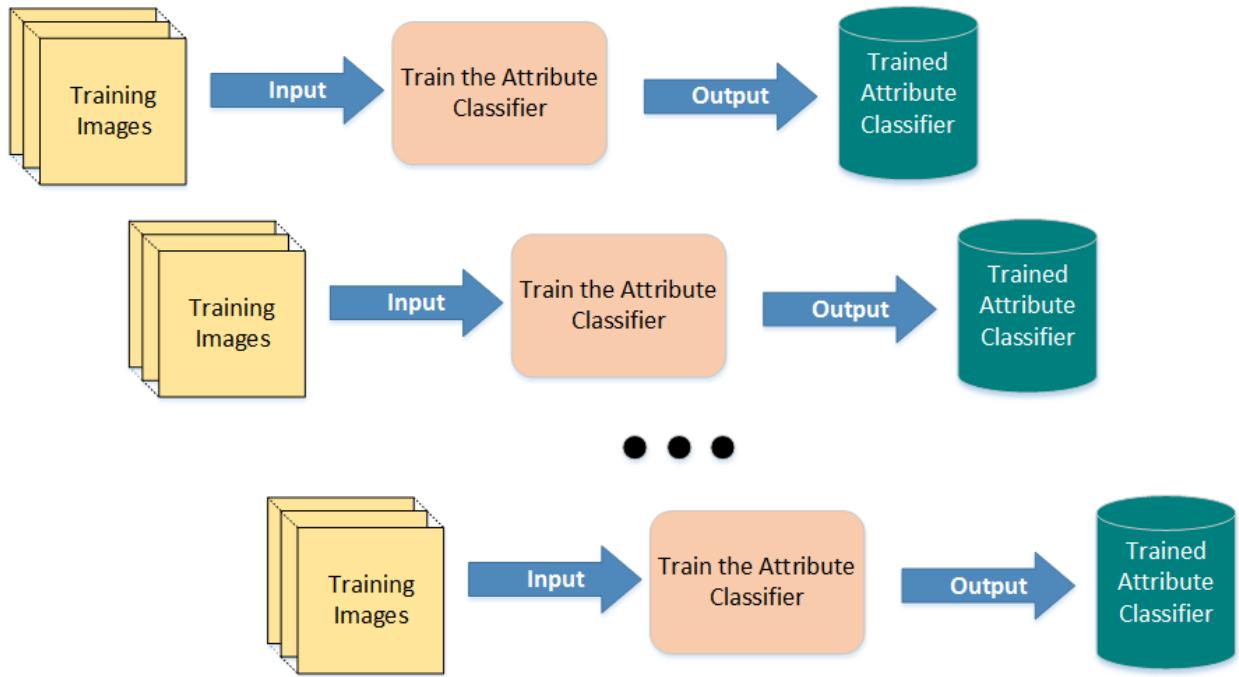
Figure 1.7: Each of the attribute classifiers are trained separately.
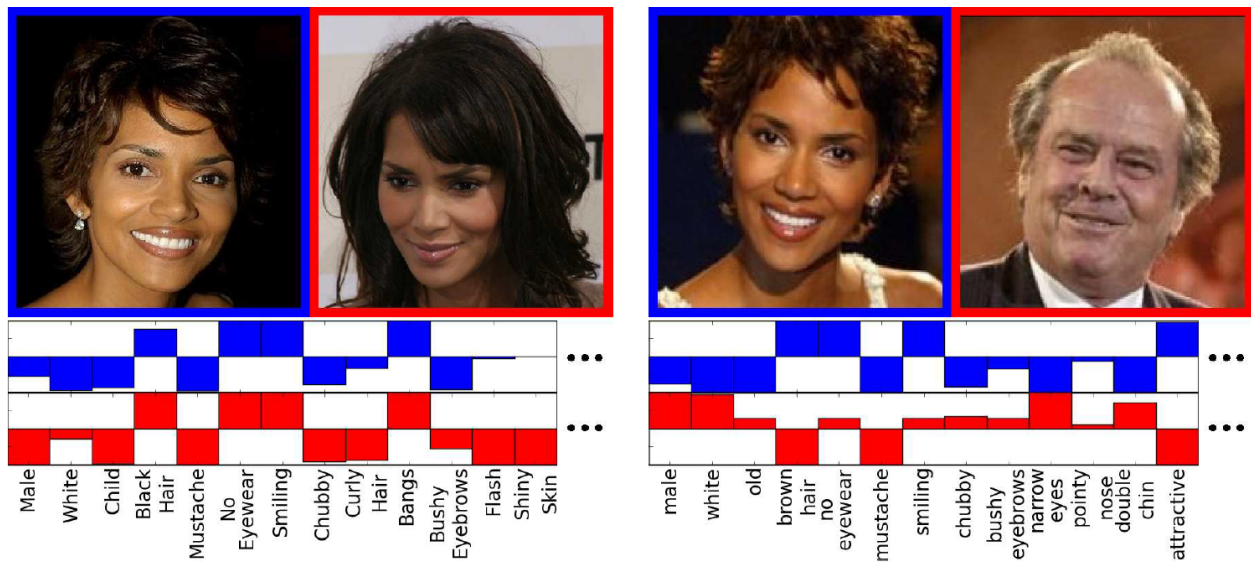
Figure 1.8: Various attributes are predicted from each face image along with a measure of how prominent each attribute is. The pair of images on the left contain the same subject while the pair on the right contain different subjects. The diagram below the images visually displays the prominence of each feature using a bar graph. The bar graph is located above the center line if the image has the attribute and below the center line if it does not have that attribute. The blue bar graph is for the image outlined in blue while the red bar graph is for the image outlined in red. Figure from [45].
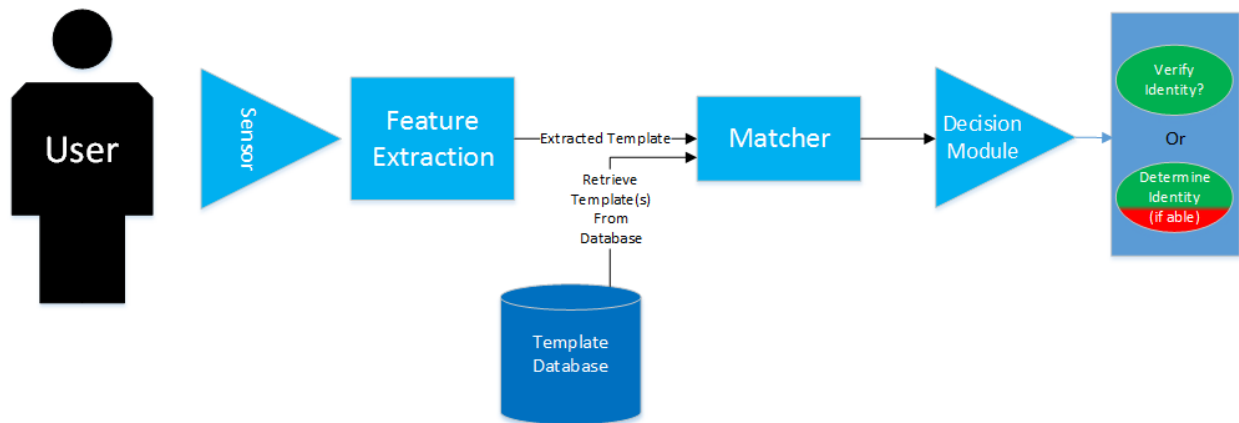
Figure 1.9: An image-based biometric system: The user presents themselves to the sensor which captures an image of a biometric trait. Features are extracted from the captured image and a feature set is generated. The generated feature set is compared to a previously generated template(s) from the template database, and a match score is generated. The decision module either verifies the claimed identity or determines which identity (if any).

be sufficient to uniquely identify a person, although attribute-based face recognition has seen some success [70]. In the field of *biometrics* [37], physical traits such as face, fingerprint or iris can be used for recognizing a person. A person may be uniquely recognized by extracting features from an image of the specified trait. The extracted features are used to generate a biometric template that is unique to that individual. Once generated, these features can be compared against previously generated templates. If a similarity score is above a configurable threshold, a match is declared.

A biometric recognition system typically acquires an image of a physical trait[1] via a sensor. The image is processed and features are extracted (see Figure 1.9). There can also be biometric systems that do not rely on an image, such as a voice-based recognition system that utilizes acoustic signals. The focus of a recognition system is in recognizing an individual; however, the captured image can also be used to predict attributes about the individual. Some attributes that may be predicted could be gender, age or race of the subject.

There are several benefits to attribute prediction in the domain of biometrics as listed in [35, 15, 38]:

- **Improve recognition rates:** Improve the recognition accuracy by fusing attribute information

---

[1]Behavioral traits such as gait, signature, or keystroke dynamics can also be used.

6

with a biometric trait [35].

- **Human understandable interpretation**: A semantic description can be generated from the biometric image. The semantic description is useful for describing the person in a human understandable language [15].

- **Robustness to low data quality:** Certain attributes may still be accurately extracted even though generation of a biometric template may not be possible due to the low quality of data [15].

- **Privacy and consent free acquisition:** Some attributes can potentially be captured from uncooperative subjects without consent [15] which may constitute a privacy violation. An 'attribute-only' approach could eliminate the need for determining the identity of an individual while still providing use for a commercial system (e.g., targeted advertising).

- **Privacy:** The identity of the individual does not have to be stored or 'identified' while the attributes may still provide useful. Targeted advertisements would be an example [15]

- **Search space reduction:** If an attribute can be predicted reliably, only a subset of the database needs to be searched. For example, if the user's gender attribute was female, only the females in the database would need to be searched [15].

- **Age specific access control:** Young children could be prevented from viewing age restricted tv shows or media [15].

- **Human Computer Interaction:** A personalized avatar could be generated based on attributes gleaned from the user's image [15].

- **Cross spectral implications:** Soft biometric attributes can potentially enable cross-spectral recognition, when images acquired in the near-infrared spectrum have to be compared against their visible spectrum counterparts [38].

Table 1.1: Examples of attribute prediction using different biometric traits.

Trait	Attribute	Method Used	Dataset (#images/#subjects)	Prediction Accuracy	Reference
Body	Gender	Figure Sequential with SVM	HumanID (100 Subjects)	96.7%	[93]
Face	Age	LBP, HOG, Bio Inspired Features with a nonlinear SVM	YGA (8000 images)	94.9%	[28]
NIR Face	Gender	LBP with SVM	CBSR NIR (3200 images)	93.59%	[67]
Fingerprint	Gender	Discrete Wavelet Transform, Wavelet Analysis	Private (498 images)	96.59%	[54]
Face	Ethnicity	2D and 3D Multi Scale Multi Ratio LBP with Adaboost	FRGCv2.0 (180 subjects)	99.5%	[94]

The prediction of attributes from biometric data has seen a lot of success specifically with the face trait. Gender, age, and race are some examples of attributes that have been successfully predicted from the face modality. In 1990, one of the first papers on predicting gender from face was published by by Golomb et al. [26]. They used a neural network to predict gender from a private database of 90 images with a 91.9% prediction accuracy [26]. Numerous papers have since been published utilizing a variety of feature descriptors, including Local Binary Patterns (LBP) [75, 29, 63], Scale-Invariant Feature Transform (SIFT) [87, 90], Histogram of Oriented Gradients (HOG) [29], raw pixels [41, 4] and deep learning methods [91].

The prediction of age from face has also seen a lot of success in the research literature. According to Dantcheva et al. [15], age from face falls into 5 main categories: a) geometric-based approaches, b) appearance-based approaches, c) aging pattern subspaces, d) age manifolds, e) and automated age classification or regression. The success of an age estimation system can be measured by 'mean absolute error' (MAE). Guo et al. [30] were able to obtain a MAE of 2.6 years on the YGA database which contains over 8,000 images.

Race is another attribute that has been assessed from facial images. As with age, predicting race also utilizes appearance and geometric based approaches. In addition, chromaticity (skin tone) as well as approaches based on local and global features have been published. Zhang and Wang [50] were able to achieve a 99.5% prediction accuracy using a subset of the FRGC v2.0 dataset with 180 subjects while Ding et al. [19] were able to achieve a 98% accuracy using a larger subset of FRGC v2.0 that contained 466 subjects.

(a) Ocular image	(b) Segmented iris region	(c) Normalized Iris

Figure 1.10: The process of iris recognition typically involves (a) imaging the ocular region of the eye using an NIR camera, (b) segmenting the annular iris region from the ocular image, and (c) unwrapping the annular iris region into a fixed-size rectangular entity referred to as a normalized iris. Image (a) is from [21].
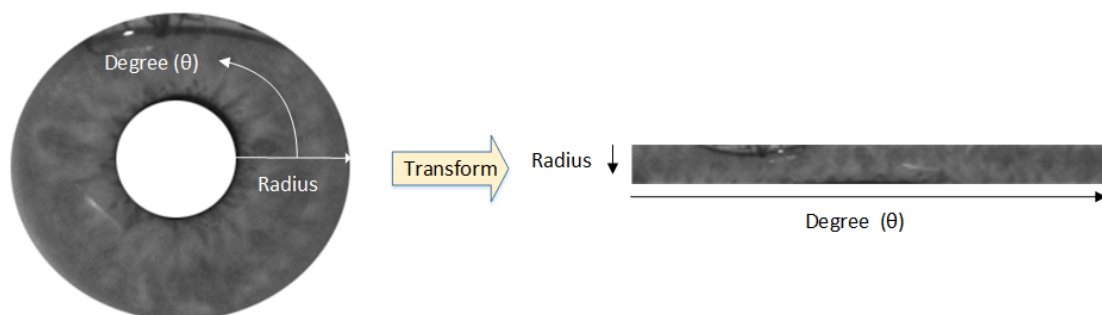
Figure 1.11: Normalization process: 'unrolling' the annular iris image by transforming it to a rectangular shape. Original iris image from [21].

1.3 Iris recognition and attribute prediction

Iris recognition systems are a type of biometric system that utilize the iris patterns evident in the eye for automated recognition of individuals [37]. The rich texture of dark colored irides is not easily discernible in the visible wavelength; therefore, the iris is typically imaged in the Near Infrared (NIR) spectrum since longer wavelengths tend to penetrate deeper into the multi-layered iris structure thereby eliciting the texture of even dark-colored eyes. Further, the NIR image acquisition process does not excite the pupil, thereby ensuring that the iris texture is not unduly

9
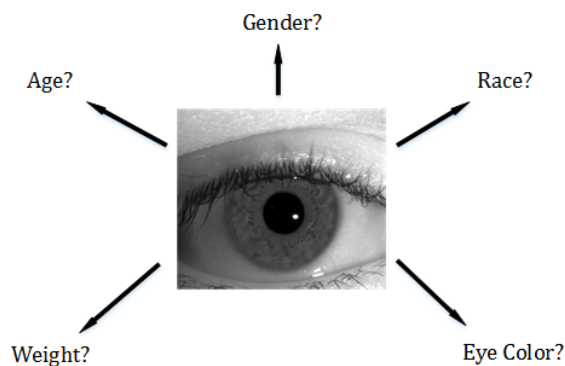
Figure 1.12: What attributes of an individual can be predicted from an NIR ocular image?

deformed due to pupil dynamics [13]. Once the image has been captured, a typical iris recognition system will segment the iris portion of the ocular image (see Figure 1.10). The annular shaped iris is 'unwrapped' into a rectangular shape by converting the cartesian coordinates of annular shape into polar coordinates, where the width of the rectangle corresponds to information in the radial direction and the height to the angular direction (see Figure 1.11). Unrolling of the iris image allows for an equal comparison between eyes with varying pupil sizes as well as the capability to utilize a fixed size image for ease of comparison.

Predicting attributes from a biometric trait such as the face has been extensively studied (See Table 1.1), while predicting attributes from the iris is a relatively less studied topic. What attributes may be predicted from the iris (see Figure 1.12)? Current iris attribute research has predicted primarily gender and race ([86, 5, 85, 8, 43]) from the iris utilizing mainly texture descriptors and more recently deep learning models ([80, 72]). Some sample NIR ocular images with the gender and race label are displayed in Figure 1.13. In addition to gender and race, the prediction of eye color as an attribute has recently been published [9] and will be presented in Chapter 3. Some sample NIR and VIS ocular images with the eye color label are displayed in Figure 1.14.

1.4 Ocular Anatomy

A review of the ocular anatomy is useful in understanding the type of gender and race markers present in the ocular region (eye color markers will be discussed in section 3.2). The ocular region could be defined as the region housing the eye (see Figure 1.15). The eyeball has both upper and
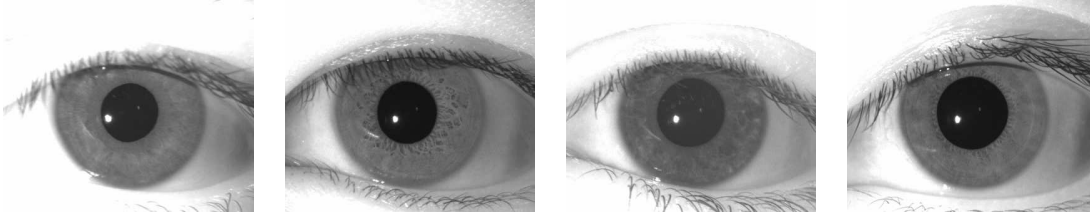
Figure 1.13: Examples of ocular images pertaining to different categories of individuals. From Left to Right: male Caucasian, male non-Caucasian, female Caucasian, female non-Caucasian. The images are from [21].
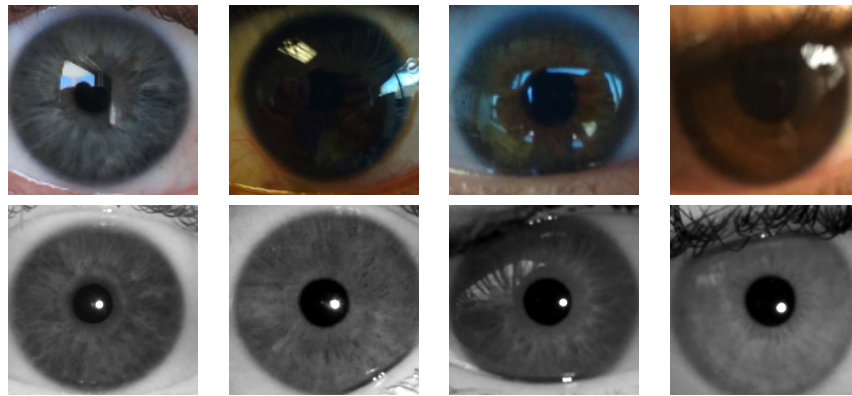
Figure 1.14: Sample eye images captured in the NIR and RGB color space demonstrating eye color as an attribute of an NIR iris image. The NIR images were taken with the Iritech IrisShield USB sensor while the RGB images were taken with a mobile phone camera.

lower eyelids that provide a protective and lubricative function to the eyeball. The upper eyelid contains the levator palpebrae superioris, which is the muscle that allows the eye to blink [3]. The gap between the upper and lower eyelid is the palpebral fissure. The iris and pupil region are located between the upper and lower eyelids.

Previous research has established the distinctiveness of the iris patterns of an individual [17]. The iris texture is imparted by an agglomeration of several anatomical features: Fuchs' Crypts, Wolfflin nodules, pigmentation dots, and contraction furrows. The iris also contains 3 different circular regions: Collarette, Ciliary Zone, and the Pupillary Ruff. There are some correlations between features that are present in the iris. For example, an iris that has no Fuchs' crypts may have clearly distinguishable contraction furrows [74]. A decrease in the density of the stroma occurs as the number of Fuchs' crypts increases. As the density decreases, the contraction furrows will also

11
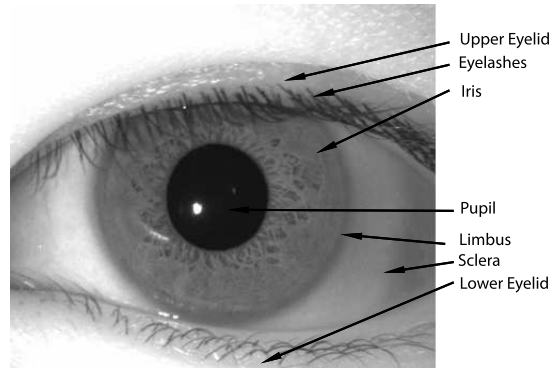
Figure 1.15: An NIR image depicting the various parts of the ocular region.

decrease.

The medical literature suggests both geometric and textural difference between male and female irides. From a textural perspective, Larsson and Pedersen [48] found that males have a greater number of Fuchs' crypts than females. From a geometric perspective, Sanchis et. al [68] report that the pupil diameters are greater in emmetropic females.[2]

If the entire ocular region is considered (and not just the iris), it has been found that the lacrimal glands of men are 30% larger and contain 45% more cells than those of females [89]. There are also significant corneal differences in such that women have steeper corneas[3] than men and their corneas are also thinner [76][89]. Other differences in the cornea include "diameter, curvature, thickness, sensitivity and wetting time of the cornea" [76]. While a number of the aforementioned features may not manifest themselves in a 2D NIR ocular image, we hypothesize that the texture of the ocular region, including the iris, may offer gender (or sex) cues of an individual.

There also exists textural differences in the iris between races as well. Edwards et al. [22] examined images of irides in the visible spectrum from 3 separate populations: South Asian, East

---

[2]The emmetropic state of the subjects in each of the datasets used througout this dissertation is unknown. An experiment was conducted to determine if there was a statistically significant difference in pupil diameter between males and females in the dataset that was used. Using the diameter of the iris as determined by a Commercial off-the-shelf (COTS) software, there was no statistically significant gender-specific difference found in either the left (male $\{\mu = 80.66, \sigma = 16.2\}$, female $\{\mu = 80.9, \sigma = 16.3\}$) or right (male $\{\mu = 79.6, \sigma = 15.5\}$, female $\{\mu = 79.7, \sigma = 16.2\}$) eyes.

[3]If we liken a cornea to a sine wave, we can think of a steep cornea as a sine wave with a higher amplitude.

12

Asian and European. Europeans were found to have a higher grade[4] of Fuchs' crypts, more pigment spots, more extended contraction furrows, and more extended Wolfflin nodules than East Asians [22]. East Asians were found to have a lower grade of Fuchs' crypts than both Europeans and South Asians. Europeans had the largest iris width, followed by South Asians, and then by East Asians [22]. As for eye color, East Asians had the darkest while Europeans had the lightest.

1.5 Dissertation Contributions

The contributions of this dissertation can be summarized below:

- **State of the art method for gender, race and eye color prediction from NIR ocular images:** A state of the art method utilizing Binarized Statistical Image Features (BSIF) as a texture descriptor and a Support Vector Machine (SVM) as a classifier, that does *not* require segmentation or normalization of the NIR ocular image. A Convolutional Neural Network architecture is also presented that yields a comparable prediction accuracy.

- **Demonstrate the ocular region provides greater gender prediction accuracy than the iris-only region:** The proposed method utilizing the texture descriptor BSIF demonstrates greater prediction accuracy for gender from NIR ocular images than from NIR iris-only images.

- **Impact of gender/race/eye color on gender/race/eye color attribute predictions from NIR ocular images:** Analysis of the impact covariates have on attribute prediction.

- **Impact of image blur on gender/race attribute prediction from NIR ocular images:** When using BSIF for texture feature extraction, the race prediction accuracy decreases at a much faster rate than that of gender as the level of image blur increases.

- **Impact of image resolution on attribute prediction from NIR ocular images:** Utilizing a Convolutional Neural Network trained on downsampled NIR ocular images, it was shown

---

[4]In their work, the authors defined 4 categories of Fuchs' crypts. Category 1 contains no crypts, while category 4 contains 'at least three large crypts located in three or more quadrants of the iris' [22].

that 5×6 images could provide an attribute prediction accuracy similar to that of much larger image resolutions, such as 340×400.

- **Introduction of a feature vector for attribute prediction:** Introduction of a feature vector entitled 'attributeCode', specifically raceCode and genderCode. An attributeCode is a feature vector generated utilizing a model trained to predict a specific attribute but is capable of predicting another attribute with only a slight decrease in prediction accuracy.

- **Raise the level of awareness on the importance of subject disjoint train/test protocols in the iris attribute prediction research community:** Brought to the forefront the importance of using a subject disjoint train and test protocol. If a subject disjoint train/test protocol is not followed, an optimistically biased classifier results. Prior to my first publication [8], the research in this field did not consistently follow a subject disjoint train/test protocol.

Currently our work has shown success in predicting race, gender and eye color from Near Infrared ocular images. Those methods will be presented in the later chapters of this proposal as well as their covariate analysis. The organization of this thesis is below:

In Chapter 2, a method to predict *gender* and *race* utilizing texture descriptors and a convolutional neural network will be presented. The NIR ocular image will be segmented into different regions and the prediction accuracy from each region will be analyzed. In addition, the impact of race, eye color, and image blur on gender and race prediction will each be analyzed.

In Chapter 3, a method to predict *eye color* utilizing texture and raw pixel intensity will be presented and contrasted.

In Chapter 4, the impact of image resolution on attribute prediction will be analyzed.

In Chapter 5, the possibility of cross-attribute prediction will be studied.

In Chapter 6, a summary of the work will be presented as well as the contributions of this dissertation.

CHAPTER 2

PREDICTING RACE AND GENDER FROM NEAR INFRARED IRIS AND OCULAR IMAGES

The majority of the work in this chapter has been published in the journal IEEE Access

2.1 Introduction

(a) (Extended) Ocular Image   (b) Iris-Only Image   (c) Iris-Excluded Ocular Image
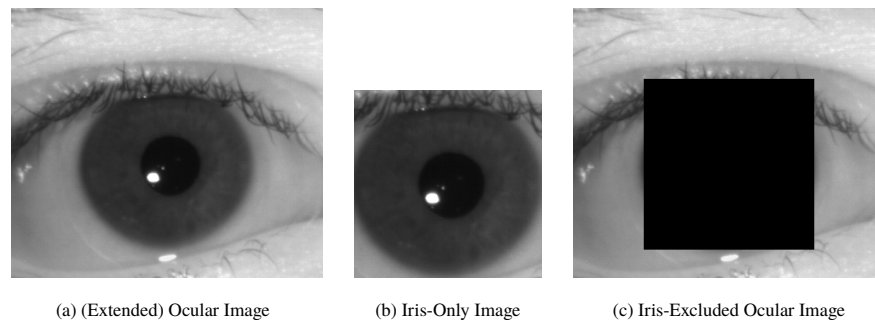
Figure 2.1: The three different regions in an NIR ocular image that are independently considered for the gender and race prediction tasks. Images are from [21].

The work discussed in this chapter will focus specifically on the prediction of race[1] and gender[2] from NIR ocular images used in iris recognition systems. Some sample NIR ocular images with the gender and race label are displayed in Figure 2.2. The images considered in this dissertation are frontal, not off-axis images. There are methods [69] for iris recognition utilizing off-axis images (up to 70 degrees), but off-axis images are not considered for this dissertation.

Previous research on this topic has extracted and used only the iris region, while most operational iris biometric systems typically acquire the extended ocular region for processing. Therefore, we investigate the gender and race predictive accuracy associated with three different regions: (a)

---

[1]The terms 'ethnicity' and 'race' have been used interchangeably in related biometric literature. An exact definition of either of these two terms appears to be debatable, and further information can be found in [7].

[2]The terms 'gender' and 'sex' have been used interchangeably in the biometric literature. There is, however, a specific definition provided by the Health and Medicine Division of the National Academies of Science, Engineering and Medicine. They state that sex is biologically or genetically determined, while gender is culturally determined [88].
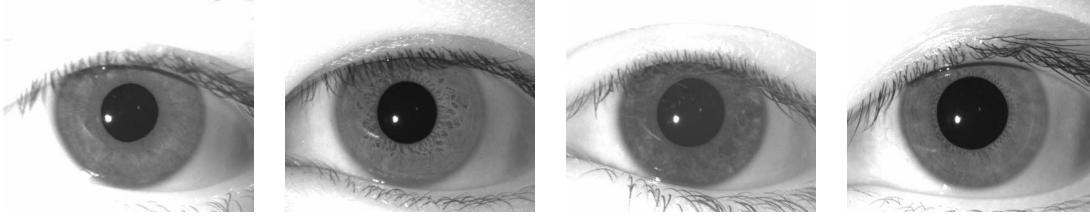
Figure 2.2: Examples of ocular images pertaining to different categories of individuals. From Left to Right: male Caucasian, male non-Caucasian, female Caucasian, female non-Caucasian. The images are from [21].

the extended ocular region; (b) the iris-excluded ocular region and (c) the iris-only region. The normalized[3] version of the iris-only region is also investigated in Section 2.6.12 for its gender prediction accuracy. See Figure 2.1 for a sample image displaying each region. We employ 3 distinct texture operators to extract features from each of these regions: BSIF (Binarized Statistical Image Feature), LBP (Local Binary Pattern) and LPQ (Local Phase Quanitzation). A Support Vector Machine (SVM) is trained as a binary classifier for each attribute (race and gender). In addition to the methods utilizing a texture descriptor, a method using a simple CNN is presented in Section 2.6.11.

In this chapter, we offer the following contributions:

- A method to predict gender and race from a NIR ocular image that is competitive with the state of the art

- The prediction accuracies due to the iris-only region is contrasted with that of the extended ocular region using multiple texture descriptors

- Demonstrate that eye laterality does not have a significant impact on gender and race prediction.

- Determine whether Caucasian or non-Caucasian subjects exhibit higher gender prediction accuracy

---

[3]The normalized iris is a rectangular rendition of the annular iris and is obtained by sampling the segmented iris region in the radial and angular directions using a rubber sheet model [18].

16

- Determine whether male or female subjects exhibit higher race prediction accuracy

- Study the sensitivity of gender and race prediction to image blur.

- Study the impact of eye color on race and gender prediction.

- Examine the normalized iris image in contrast to the non-normalized iris image for gender prediction

2.2 Related Work - Gender Prediction

One of the earliest work in prediction of sex from the iris was published by Thomas et al. [86]. The authors assembled a dataset of $57,137$ ocular images. The iris was extracted from each of the ocular images and normalized into a 20×240 image using Daugman's rubber sheet method [17]. A feature vector was generated from the normalized image by applying a one dimensional Gabor filter. Feature selection was performed using an information gain metric. The resulting reduced feature vector was classified as 'Male' or 'Female' by a decision tree algorithm. Only left iris images were used for their experiments. The authors were able to achieve 'upwards of 80% with Bagging' [4] using only the Caucasian subjects in their dataset.[5]

Bansal et al. [5] were able to achieve an 83.06% sex classification accuracy using statistical and wavelet features along with an SVM classifier. Occlusions from the iris region were removed (i.e., eyelids, eyelashes) using a masking algorithm. The size of their dataset, however, was quite small with only 150 subjects and 300 iris images. 100 of the subjects were male and 50 of the subjects were female. However, it is not clear if they used a subject-disjoint evaluation protocol.

Lagree and Bowyer performed sex classification on a dataset of 600 iris images each of which was normalized to a 40×240 rectangular image. 8 horizontal regions of 5-pixel width and 10 vertical regions of 24-pixel width were then created. Using the created regions and some simple texture filters (i.e., for detecting spots, lines), an 882-dimensional feature vector was computed.

---

[4]It was not stated whether a subject-disjoint training and test set were used.

[5]A prediction accuracy of 75% was achieved using all of the images.

An SVM classifier was then applied (specifically the WEKA SMO algorithm) for classification, achieving a 62% accuracy.

Tapia et al. [80] continued their work on deducing gender from iris by utilizing a CNN architecture that fused *normalized* iris images from the left and right eye and were able to achieve a 84.66% accuracy. Tapia et al. [80] cited the small size of their dataset as a possible reason for their performance not surpassing that of their previous work [85]. In [81], Tapia & Aravena proposed a CNN architecture that fused the *periocular* NIR images together. The model utilized 3 CNNs: one for the left eye, one for the right eye and one to fuse the left and right eye models together. They were able to achieve an 87.26% prediction accuracy.

Fairhurst et al. [14] utilized geometric features from the ocular image and texture features from the normalized iris image and were able to achieve an 81.43% prediction accuracy on a subset of the BioSecure dataset consisting of 200 subjects and 1600 images.

Singh et al. [73] use a variant of an auto-encoder that includes the attribute class label along side of the reconstruction layer. They used NIR ocular images that were resized to 48×64 pixels. Their proposed method was tested on both the ND-GFI and ND-Iris-0405 datasets from Notre Dame. The experiments on the ND-GFI dataset utilized the 80-20 subject-disjoint split specified in the dataset. While experiments using the ND-Iris-0405 dataset were not indicated as being subject-disjoint, their paper states: 'All protocols ensure mutually exclusive training and testing sets, such that there is no *image* [emphasis added] which occurs in both the partitions'.

Kuehlkamp et al. [43] studied the effect of mascara on predicting gender from iris. Using only the occlusion mask from each of the images, they achieved a 60% gender prediction accuracy. They went on to show that LBP combined with an MLP network was able to achieve a 66% accuracy. Using the entire 'eye' image they were able to achieve around 80% using CNNs and MLPs. In 2019, Kuehlkamp et al. [44] published a study that confirmed the findings of Bobeldyk and Ross [8], the periocular region (and occlusions of the iris) contribute more to the sex predictive accuracy than the iris texture alone [6].

---

[6]It should be noted that just prior to publication of this document, Tapia et al. [83] published a paper claiming a prediction accuracy of 93.45% for left and 95.45% from the *normalized iris*

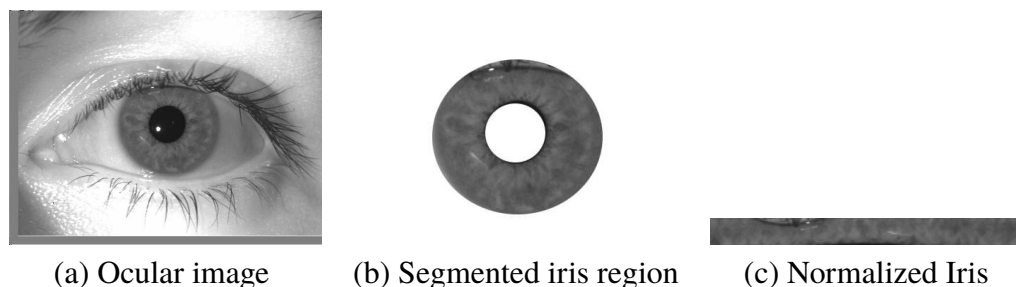|		
(a) Ocular image	(b) Segmented iris region	(c) Normalized Iris
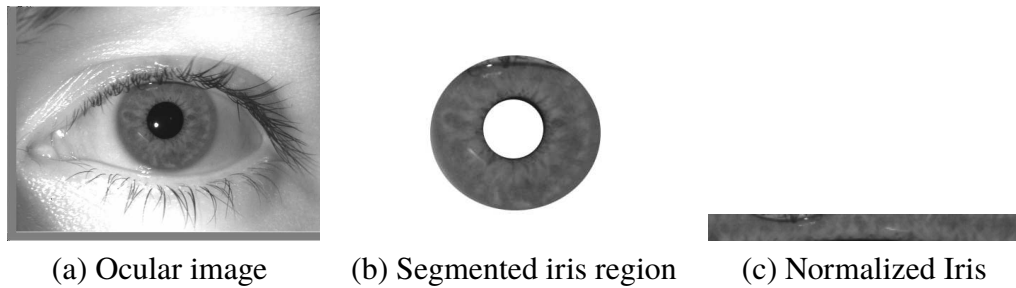
Figure 2.3: The process of iris recognition typically involves (a) imaging the ocular region of the eye using an NIR camera, (b) segmenting the annular iris region from the ocular image, and (c) unwrapping the annular iris region into a fixed-size rectangular entity referred to as normalized iris. Image (a) is from [21].

Tapia et al. [81] used a feature selection model that was similar to their earlier work on iris [85], but applied it to periocular images. They were able to achieve a 90.0% prediction accuracy on a dataset containing 120 subjects and 1920 images.

Previous work utilizing a *single* eye image (left or right) are displayed in Table 2.1 and those utilizing a *fused* model combining the left and right are shown in 2.2. For ease of comparison, the works that utilized the publicly available gender labeled ND-GFI dataset are highlighted in Table 2.1.

Most biometric recognition work pertaining to NIR iris images have focused on extracting the iris region from the captured ocular image (see Figure 2.3). Thus, algorithms for soft biometric prediction have typically focused on the iris region rather than the extended ocular region (see Figure 2.1). Predicting soft biometric attributes from the ocular region provides one major advantage over the iris region in that it does not require a potentially error prone algorithm for iris region extraction. In this chapter, a method will be proposed and the sex prediction accuracy will be examined to determine which region provides the greater sex cues . Section 2.6.12 will include analysis of the normalized iris region (see Figure 2.16).

It should also be noted that there are some sex prediction work using the periocular region in the *visible wavelength spectrum* in [51], [65], [82] and [52].

---

region only. These results appear to be in direct contrast with the findings from our research lab and the researchers at Notre Dame.

Table 2.1: Gender Prediction - Related Work (Left or Right Eye Image). Work that uses the publicly available ND-GFI dataset are highlighted for ease of comparison.

Authors	Year	Subject-Disjoint Specified	Dataset	Number of Subjects	Number of Images	Features	Prediction Accuracy
Thomas et al. [86]	2007	No	Private	Unknown	57,137	Geometric/Texture features	80%
Bansal et al. [5]	2012	No	Private	150	300	Statistical/Texture features	83.06%
Singh et al. [73]	2017	No	ND-Iris-0405	356	60,259	Deep class-encoder	82.53%
Singh et al. [73]	2017	Yes	ND-GFI	1500	3000	Deep class-encoder	83.17%
Lagree & Bowyer [46]	2011	Yes	Private	120	1200	Basic texture filters	62%
Fairhurst et al. [14]	2015	Yes	BioSecure	200	1600	Geometric and Texture Features	81.43%
Bobeldyk & Ross [8]	2016	Yes	Private	1083	3314	BSIF	85.7 %
Kuehlkamp et al. [43]	2017	Yes	ND-GFI	1500	3000	CNN and MLPs	80%
Tapia et al. [82]	2017	Yes	CROSS-EYED	120	1920	HOG w/ feature selection	90.0%
This Work	2018	Yes	ND-GFI	1500	3000	BSIF	84.4%
This Work	2018	Yes	BioCOP2009	1096	41,780	BSIF	86.0%

Table 2.2: Gender Prediction - Related Work (Left + Right Eye Fusion).

Authors	Year	Subject-Disjoint Specified	Dataset	Number of Subjects	Number of Images	Features	Prediction Accuracy
Tapia et al. [84]	2014	No	ND-GFI	1500[7]	3000	Uniform LBP	91.33%
Tapia et al. [85]	2016	Yes	ND-GFI	1500	3000	Iriscode and weighted feature selection	89%
Tapia et al. [80]	2017	Yes	ND-GFI	1500	3000	CNN fusing of separate left/right CNNs	84.66%
Tapia and Aravena [81]	2018	Yes	ND-GFI	1500	3000	CNN (Reduced version of LeNet)	87.26%

2.3   Related Work - Race Prediction

The problem of attribute prediction, is typically posed as a pattern classification problem where a feature set extracted from the biometric data (e.g., an ocular image) is input to a classifier (e.g., SVM, decision tree, etc.) in order to produce the attribute label (e.g., 'Caucasian'). The classifier itself is trained in a supervised manner with a training set consisting of ocular data labeled with attributes. The performance of the prediction algorithm is then evaluated on an independent test set. Good practice [36] dictates that the subjects in the training set and test set are mutually exclusive. An optimally biased predictor can be produced if there is an overlap of subjects in the training and test sets as indicated in [8, 85]. While most recent work in attribute prediction from iris [8, 85] have clearly adopted a subject-disjoint protocol, some of the earlier papers on this topic have been ambiguous on this [86, 5, 60, 61]. Table 2.3 summarizes the previous work on race prediction.

There are only a few papers that attempt to deduce race from NIR iris images. In [60] and [61]

---

[7]The published paper claims 1500 subjects; however it was discovered during our experiments that there was actually far less number of subjects. The authors confirmed this error via email and in one of their subsequent publications [85].

Table 2.3: Race Prediction - Related Work.

Authors	Subject Disjoint Specified	Dataset Used	# of subjects	# of images	Features Used	Prediction Accuracy
Qiu et al.[60]	No	CASIA, UPOL, UBIRIS	Unknown	3982	Gabor filters	85.95%
Qiu et al.[61]	No	Proprietary	60	2400	Gabor filters	91.02%
Singh et al.[73]	No	ND-Iris-0405/Multi-Ethnicity	240/Unknown	60,259/60,310	Deep class-encoder	94.33%/97.38%
Lagree and Bowyer[46]	Yes	Proprietary	120	1200	Basic texture filters	90.58%
Proposed Work[10]	Yes	BioCOP2009	1096	41780	BSIF	90.1%

the authors do not state whether their train and test partitions are subject-disjoint, and the size of the datasets are quite small (3982 and 2400 images, respectively). In both publications, Qiu et al. [60, 61] utilized the texture generated from Gabor filters to create a feature vector that was classified using AdaBoost and SVM (respectively) classifiers. A smaller region of the captured iris image was used in order to minimize occlusions from eyelids or eyelashes.

Singh et al. [73] also did not specify a subject-disjoint experimental protocol. Their proposed method used a variant of an auto-encoder that includes the class label along side of the reconstruction layer. The experiments were performed on the ND-Iris-0405 dataset as well as a multi-ethnicity iris dataset composed of 3 separate datasets. Each class (Asian, Indian, Caucasian) was represented by a distinct dataset. They achieved a 94.33% prediction accuracy on the ND-Iris-0405 dataset and 97.38% on the multi-ethnicity iris dataset. However, it is not clear if the multi-ethnicity results were optimistically biased due to the use of different datasets for the 3 classes. As pointed out by El Naggar and Ross [23], dataset-specific cues are often present in the images.

2.4 Feature Extraction

One of the goals of our work is to establish the utility of simple texture descriptors for attribute prediction. Uniform local binary patterns (LBP) [58] and binarized statistical image features (BSIF) are two texture descriptors that have performed well on the Outex and Curet texture datasets [40]. Both have shown to perform well in the attribute prediction domain [8, 84], with BSIF outperforming LBP in both domains (texture and attribute prediction). Three texture descriptors were considered in this chapter: BSIF, LBP and LPQ (Local Phase Quantization).

LBP [58] encodes local texture information by comparing the value of every pixel of an image with each of its respective neighboring pixels. This results in a binary code whose length is equal
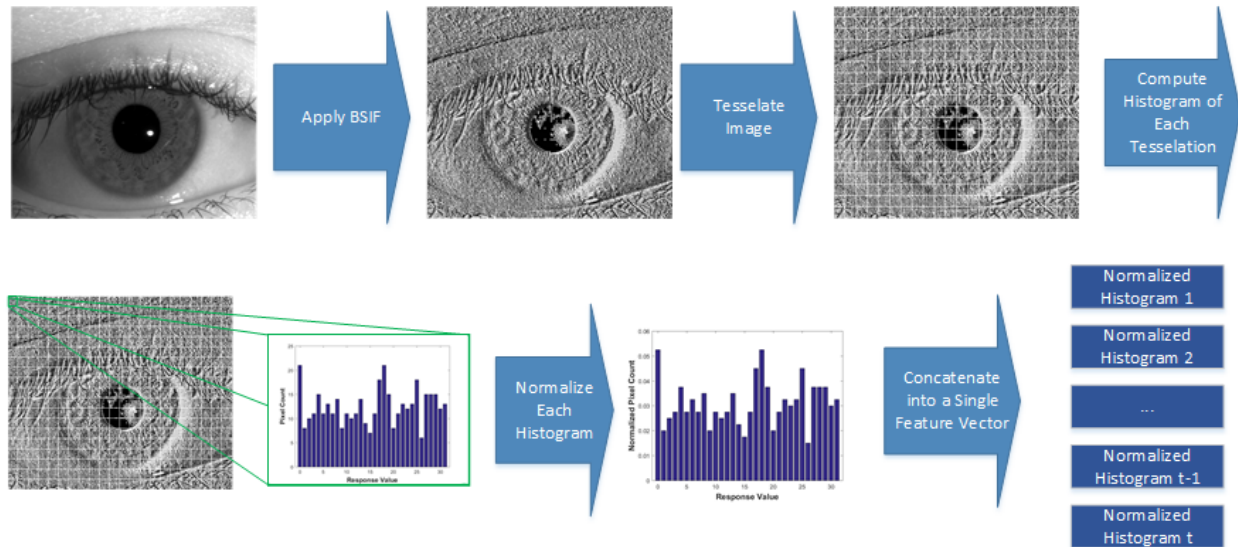
Figure 2.4: Generating the feature vector for gender and race prediction classification based on BSIF. A bit size of 8 and filter size of 3×3 was used as the BSIF parameters in this illustration.

to the number of neighboring pixels considered. The binary sequence is then converted into a decimal value, thereby generating an LBP code for the image.

LPQ [59] encodes local texture information by utilizing the phase information of an image. A sliding rectangular window is used, so that at each pixel location, an 8 bit binary code is generated utilizing the phase information from the 2-D Discrete Fourier Transform. A histogram of those generated values results in a 256-dimensional feature vector.

BSIF was introduced by Kanala and Rahtu [40] as a texture descriptor. BSIF projects the image into a subspace by convolving the image with pregenerated filters. The pregenerated filters are created from 13 natural images supplied by the authors of *Natural Image Statistics* [34]. 50,000 patches of size $k \times k$ are randomly sampled from the 13 natural images. Principal component analysis is applied, keeping only the top **n** components. Independent component analysis is applied generating **n** filters of size $k \times k$. The authors of [40] provide the pregenerated filters for $k = \{3, 5, 7, 9, 11, 13, 15, 17\}$ and $n = \{5 - 12\}$.[8]

Each of the **n** pregenerated filters are convolved with the image and the response is binarized. If the response is greater than zero, a '1' is generated. If the response is less than or equal to

_____

[8]For $n = \{9 - 12\}$, $k = 3$ was not made available by [40].

22

|(a) Ocular Image|(b) Iris-Only Image|(c) Iris-Excluded Ocular Image|
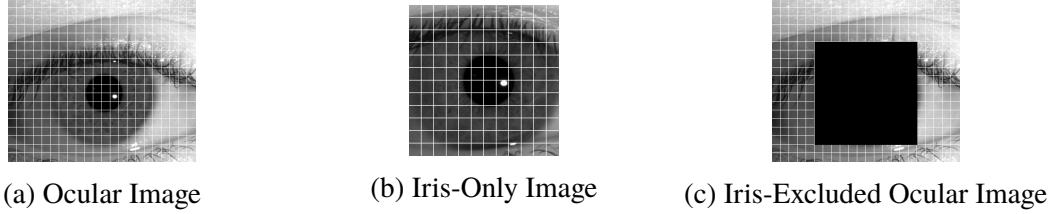
Figure 2.5: Tessellations applied to the three image regions. The images are from [21].

Table 2.4: Dataset summary including attribute labels used for each dataset.

Dataset	Available Images	Number of Subjects	Attribute
BioCOP2008	3,314	1,083	Gender
BioCOP2009-1	41,831	1,096	Race, Gender, Eye Color
BioCOP2009-2	40,394	1,028	Race, Gender, Eye Color
BioCOP2009-3	43,454	1,096	Race, Gender, Eye Color
BioCOP2009-4	43,281	1,028	Race, Gender, Eye Color
BioCOP2009-5	40,641	1,096	Race, Gender, Eye Color
ND Cosmetic Contact	3,000	175	Race, Gender
ND-GFI	1,944	324	Gender

zero, a '0' is generated. The concatenated responses form a binary string that is converted into a numeric decimal value (the BSIF code). For example, if the **n** binary responses were $\{1, 0, 0, 1, 1\}$, the resulting decimal value would be '19'. Therefore, given **n** filters, the BSIF response will range between 0 and $2^n - 1$.

Our proposed method applies the texture descriptor to each of the NIR ocular images which were then tesselated into 20×20 pixel regions (see Figure 2.5 for a visual representation). This tessellation was done in order to ensure the spatial information is included in the feature vector that is being created. Histograms were generated for each of the tessellations, normalized, and concatenated into a single feature vector. In order to provide consistent spatial information across each image, a geometric alignment was applied to the original NIR ocular image. The parameters chosen for this geometric alignment are similar to those proposed by [8] and discussed in Section 2.5.1 as well as shown in Figure 2.6. A graphic illustrating the feature vector generation process is shown in Figure 2.4.

2.5 Datasets

Four separate datasets were used to conduct the experiments in this chapter. The largest of the 4 datasets is the BioCOP2009 dataset, which is described in Section 2.5.1. The BioCOP2009 dataset was preprocessed based on 5 different requirements and used for the experiments throughout this chapter, each of those subsets are described in Sections 2.5.1.1, 2.5.1.2, 2.5.1.3, 2.5.1.4 and 2.5.1.5. Three other datasets were used for cross testing in order to demonstrate the generalizability of the proposed method. Those datasets are the Notre Dame (ND) Cosmetic Contact dataset (see Section 2.5.2), the ND-GFI dataset (see Section 2.5.3) and the BioCOP2008 dataset (see Section 2.5.4). The datasets along with the labeled attributes that were used for this chapter are displayed in Figure 2.4. There are not many datasets with this type of data labeled, availability of datasets helped determine the various categories of ethnicities used for the race prediction experiments in Chapter 2. It is also important to note that the two datasets (ND Cosmetic Contact and ND-GFI) used for cross testing were collected at an entirely different location than the BioCOP2009 dataset. The BioCOP2009 dataset was collected at West Virginia University, while the ND Cosmetic Contact and ND-GFI dataset were both collected at Notre Dame University. Cross testing on datasets collected at different locations greatly decreases the chance that they will contain the same subjects while introducing substantial variability in the images due to changes in factors such as lighting and sensors. The BioCOP2008 dataset was used for the experiments involving the normalized iris region in Section 2.5.4. The BioCOP2008 dataset was used for the earlier work before the BioCOP2009 dataset was made available. A summary of the datasets are shown in Table 2.4.

2.5.1 BioCOP2009 Dataset

The BioCOP2009 dataset contains NIR ocular images captured with 3 different sensors: the LG ICAM 4000, CrossMatch I SCAN2 and Aoptix Insight. The LG and Aoptix sensors captured NIR ocular images of size 640×480, while the CrossMatch sensor produced images of size 480×480. From the original BioCOP dataset, there are 5 subsets created. Each subset was generated based

on the specifications detailed in the following 5 subsections (Subsections 2.5.1.1, 2.5.1.2, 2.5.1.3, 2.5.1.4, 2.5.1.5).

5 random subject-disjoint partitions were created for the experiments that use each of the datasets. In each of those experiments, all of the images of a subject were used for either training or testing. Given some subjects have more images than others, the total number of training and testing images can fluctuate across the 5 random partitions. It is also important to note that the training and testing partitions contain images from all 3 sensors.

2.5.1.1 BioCOP2009-1

Using a commercially available SDK, the images were preprocessed to find the coordinates of the iris center and the radius of the iris. During the preprocessing stage, 276 images were rejected as the software was unable to automatically locate those coordinates. In order to ensure that all images are spatially aligned, the images were geometrically adjusted using the method outlined in [8]. The geometric alignment centers the image, using the coordinates computed by the commercial SDK, and rescales the image to a fixed size. Given that the CrossMatch sensor images were smaller than those in [8], all of the images were aligned to the smaller dimension size of 400×340 (as opposed to 440×380 in [8]). A diagram displaying the pixel measurements, as well as a sample geometrically aligned image, are shown in Figure 2.6. Images that did not contain sufficient border size after the geometric alignment were not used in the experiments (see Table 2.6).

There are 1096 total subjects in the post processed BioCOP2009-1 dataset, for a total of 41,830 images. The Aoptix and LG ICAM sensors have a left eye image for every subject, while the CrossMatch sensor has 106 subjects with no left eye images. For the right eye, the LG ICAM has an image for every subject, the Aoptix has 2 subjects with no images and the CrossMatch has 103 subjects with no images. A summary of the sensor breakdown is shown in Table 2.5. The subject attribute information for gender and race are listed in Tables 2.7 and 2.8, respectively.

Table 2.5: Statistics of the post processed BioCOP2009-1 dataset used in Chapter 2.

Sensor	Subjects with Left Images	Number of Left Images	Subjects with Right Images	Number of Right Images
Aoptix	1096	5449	1094	5389
CrossMatch	990	4528	993	4593
LG ICAM	1096	10,940	1096	10,931
Overall	1096	20,917	1096	20,913

Table 2.6: Statistics of the BioCOP2009-1 dataset used in Chapter 2. The first column denotes the number of images that were initially present in the BioCOP2009 dataset. The second column lists the number of images that were successfully preprocessed by the COTS SDK in order to find the coordinates of the iris center and the iris radius. The third column presents the number of images that contained sufficient border pixels after the geometric alignment step.

Sensor	Initial Number of Images	Post SDK Preprocessing	Post Geometric Alignment
LG ICAM 4000	21,940	21,912	21,871
CrossMatch I SCAN 2	10,890	10,643	9,121
Aoptix Insight	10,980	10,979	10,838
Total	43,810	43,534	41,830

Table 2.7: Gender statistics for the BioCOP2009-1 dataset used in Chapter 2.

Attribute		Subject Number	Left Images	Right Images
Gender	Male	467	9,035	9,009
	Female	629	11,882	11,904

Table 2.8: Race statistics for the BioCOP2009-1 dataset used in Chapter 2.

Attribute		Dataset Label	Subject Number	Left Images	Right Images
	Caucasian	Caucasian	836	16,068	16,019
		African	16	298	303
		African American	35	625	635
		American Indian	2	40	40
		Asian	73	1,347	1,351
Race	**Non-Caucasian**	Asian Indian	60	1,151	1,156
		Hispanic	33	609	630
		Middle Eastern	22	422	420
		Other	15	290	290
		Other Pacific Islander	2	36	39
		Unknown	2	31	30

(a) Before   (b) Geometric Alignment   (c) After
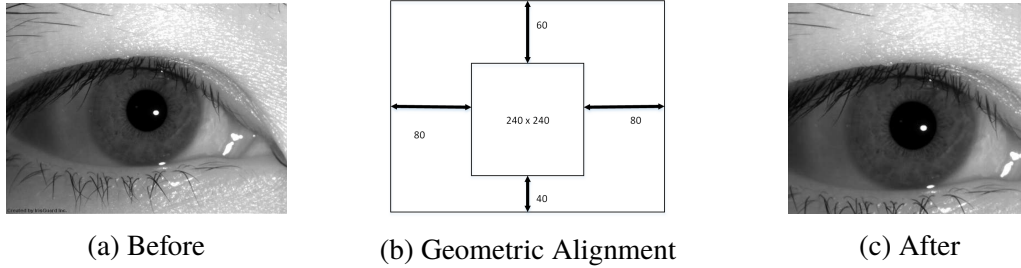
Figure 2.6: Example of a geometrically adjusted image. The geometric alignment shown was used for the BioCOP2009-1, BioCOP2009-2, BioCOP2009-4 and BioCOP2009-5 datasets. The image in (a) is from [21].

Table 2.9: Race statistics for the BioCOP2009-2 Dataset used in Chapter 2.

Attribute		Dataset Label	Subject Number	Left Images	Right Images
	Caucasian	Caucasian	781	15,061	16,019
		African	14	263	268
		African American	35	625	635
		American Indian	2	40	40
		Asian	69	1,277	1,280
Race	**Non-Caucasian**	Asian Indian	57	1,112	1,113
		Hispanic	29	543	560
		Middle Eastern	22	422	420
		Other	15	290	290
		Other Pacific Islander	2	36	39
		Unknown	2	31	30

2.5.1.2 BioCOP2009-2

The BioCOP2009-2 dataset was preprocessed the same as the BioCOP2009-1 dataset (see Section 2.5.1.1); however, there are only 247 non-Caucasian subjects and 781 Caucasian subjects that were made available for the race prediction experiments. This additional constraint was applied to allow equivalent comparison across all versions of our race experiments. The original version of the preprocessed dataset (which subsequently did not generate any good results or published experiments) contained 247 non-Caucasian and 781 Caucasian subjects. The number of subjects/images by race are listed in Table 2.9. The BioCOP2009-2 dataset is used for the race prediction experiments in Chapter 2.

Table 2.10: Summary of the BioCOP 2009-3 dataset used in Chapter 3.

Sensor Type	Number Of Images		
	Original	Post COTS SDK	Post Geometric Alignment
LG ICAM 4000	21,940	21,912	21,893
CrossMatch I SCAN 2	10,890	10,643	10,583
Aoptix Insight	10,980	10,979	10,978
Total	43,810	43,534	43,454

2.5.1.3   BioCOP2009-3

The BioCOP2009 dataset contains 43,810 NIR ocular images captured with 3 different iris sensors: LG ICAM 4000, CrossMatch I SCAN2 and Aoptix Insight. Using a commercially available SDK, the same as used in Section 2.5.1.1, the center and radius of the iris in each image were determined. During this stage, 276 images were rejected, as the software was unable to automatically locate the iris in them. To ensure spatial consistency across all the images, each image was resized to a fixed iris radius of 120 pixels, resulting in images of dimension 240×240. Images that did not include the full iris were excluded. Given there were no top, bottom, left or right border constraints as discussed in Section 2.5.1.1, a greater number of images were made available (see Table 2.10).

The BioCOP2009-3 dataset contains 6 different color labels: 'Brown', 'Blue', 'Green', 'Hazel', 'Gray' and 'Other'. The number of images pertaining to each color category is listed in Table 2.11. Category A defines the subset of images with the label 'Brown' for eye color. Category B defines the subset of images labeled as 'Blue', 'Green', 'Hazel' or 'Gray'. Images with the label 'Other' were not used in the experiments. The number of subjects included in each of these categories, as well as gender and ethnicity statistics are listed in Table 2.12. The BioCOP2009-3 dataset is used for the eye color prediction experiments in Chapter 3. No race specific dataset version was created (as BioCOP2009-2 was for BioCOP2009-1) as this dataset was used for eye color prediction experiments and not race prediction. See Chapter 3 for experiments conducted utilizing this dataset.

Table 2.11: Number of images for each color category and label of the BioCOP2009-3 dataset used in Chapter 3.

Class	Color Label	Left Eye Number of Images	Right Eye Number of Images
Category A	Brown	9862	9848
Category B	Blue	5821	5794
	Green	2825	2834
	Hazel	2699	2692
	Gray	160	160
Unknown	Other	379	380

Table 2.12: Eye Color, ethnicity and gender statistics of the BioCOP 2009-3 dataset used in Chapter 3.

Class	Eye Color	Caucasian	Non-Caucasian	Male	Female
Category A	Brown	267	228	235	260
Category B	Blue	294	2	119	177
	Green	137	6	46	97
	Hazel	130	6	50	86
	Gray	8	0	3	5
Not Used	Other	0	18	14	4

2.5.1.4 BioCOP2009-4

The BioCOP2009 dataset contains 43,810 NIR ocular images captured by 3 different iris sensors: LG ICAM 4000, CrossMatch I SCAN2 and Aoptix Insight. Using a commercially available SDK (an updated version of that used in Sections 2.5.1.1 and 2.5.1.2), the center and radius of the iris in each image were determined. During this stage, 90 images were rejected, as the software was unable to automatically locate the iris in them. To ensure spatial consistency across all the images, each image was resized to a fixed iris radius of 120 pixels, a top border of 60 pixels, a bottom border of 40 pixels and lateral borders of 80 pixels each (see Figure 2.6). The resulting image size was 340×400. Images that did not include the full iris were excluded (see Table 2.13).

The BioCOP2009 dataset contains 6 different eye color labels: 'Brown', 'Blue', 'Green', 'Hazel', 'Gray' and 'Other'. Eye color prediction is treated as a binary class problem. Category A defines the subset of images with the label 'Brown' for eye color. Category B defines the subset of images labeled as 'Blue', 'Green', 'Hazel' or 'Gray'. Images with the label 'Other' were not used in the

Table 2.13: Summary of the number of images in the BioCOP2009-4 dataset used in Chapter 4.

Sensor Type	Number Of Images		
	Original	After COTS SDK	After Crop and Resize
LG ICAM 4000	21,940	21,887	21,865
CrossMatch I SCAN 2	10,890	10,861	10,457
Aoptix Insight	10,980	10,972	10,959
Total	43,810	43,720	43,281

Table 2.14: Race statistics for the BioCOP2009-4 Dataset used in Chapter 4.

	Attribute	Dataset Label	Subject Number	Left Images	Right Images
Race	Caucasian	Caucasian	836	16,567	16,516
	Non-Caucasian	African	16	315	313
		African American	35	674	669
		American Indian	2	40	40
		Asian	69	1,424	1,419
		Asian Indian	57	1,200	1,208
		Hispanic	33	647	645
		Middle Eastern	22	433	428
		Other	15	298	298
		Other Pacific Islander	2	38	39
		Unknown	2	35	35

Table 2.15: Gender and eye color statistics of the BioCOP2009-4 dataset used in Chapter 4.

Attribute		Subject Number	Left Images	Right Images
Gender	Male	467	9,273	9,239
	Female	629	12,398	12,371
Eye Color	Brown	495	9,726	9,704
	Not Brown	583	11,571	11,527
	Other	18	374	379

experiments. Gender is viewed as a binary attribute, either 'Male' or 'Female'. Race is also treated as a binary attribute, using the labels 'Caucasian' and 'Non-Caucasian'.

The number of subjects and images for each race are listed in Table 2.14, while gender and eye color are listed in Table 2.15. The BioCOP2009-4 dataset is used for the experiments in Chapter 4.

Table 2.16: Race statistics for the BioCOP2009-5 Dataset used in Chapter 4.

Attribute		Dataset Label	Subject Number	Left Images	Right Images
	Caucasian	Caucasian	781	15,493	15,439
Race	**Non-Caucasian**	African	14	276	275
		African American	35	674	669
		American Indian	2	40	40
		Asian	69	1,347	1,342
		Asian Indian	57	1,146	1,154
		Hispanic	29	575	570
		Middle Eastern	22	433	428
		Other	15	298	298
		Other Pacific Islander	2	35	39
		Unknown	2	35	35

2.5.1.5 BioCOP2009-5

The BioCOP2009-5 dataset was preprocessed using the same methods as the BioCOP2009-4 dataset (see Section 2.5.1.4). A single difference exists, in that there are 247 non-Caucasian subjects and 781 Caucasian subjects that were made available for the race prediction experiments. This additional constraint was applied in order for equivalent comparison across all versions of the intra-dataset experiments. The original version of the preprocessed dataset (which subsequently did not generate any good results or published experiments) contained 247 non-Caucasian and 781 Caucasian subjects. The number of subjects/images by race are listed in Table 2.16.

2.5.2 ND Cosmetic Contact Dataset

In order to perform cross dataset testing, we used the Cosmetic Contact Lens dataset assembled by researchers at Notre Dame [21]. The Cosmetic Contact Lens dataset contains images that are labeled with both race and gender labels. The dataset contains images collected by 2 separate sensors, the LG4000 and the AD100. For the LG4000 sensor, 3000 images were collected for training a classifier and 1200 images were collected for testing that classifier. For the AD100 sensor, 600 images were collected for training a classifier and 300 images were collected for testing that classifier. For the purposes of our experiments we only used the LG4000 sensor images. The

Table 2.17: Gender statistics for the CCD1 Dataset used in Chapter 2.

Attribute		Subject Number	Left Images	Right Images
Gender	Male	90	950	850
	Female	85	600	600

Table 2.18: Race statistics for the CCD1 Dataset used in Chapter 2.

Attribute		Dataset Label	Subject Number	Left Images	Right Images
Race	**Caucasian**	White	105	1210	940
	Non-Caucasian	Asian	41	150	320
		Black	5	30	30
		Other	24	60	160

Table 2.19: Gender statistics for the CCD2 Dataset used in Chapter 2.

Attribute		Subject Number	Left Images	Right Images
Gender	Male	38	350	350
	Female	38	250	250

rest of the paper will refer to the 3000 images collected from the LG4000 sensor as Cosmetic Contact Dataset One (CCD1) and the 1200 verification images as Cosmetic Contact Dataset Two (CCD2).

The geometric alignment process that was used for the BioCOP2009-1 dataset (see Figure 2.7) was applied to the CCD1 and CCD2 datasets. After the geometric alignment procedure, only 4 images from CCD2 were discarded due to insufficient border size and no images were discarded from CCD1. The subject attribute information for gender and race are listed in Table 2.17 and Table 2.18 for CCD1 respectively; while the gender and race attribute information for CCD2 are listed in Table 2.19 and Table 2.20 respectively. During cross dataset testing, these 2 datasets were tested using the 5 SVM classifiers that were obtained from the 5 random partitions of the BioCOP2009 training set. Using the same SVM classifiers allows for a fair comparison between the prediction accuracies of the intra-dataset and cross-dataset test scenarios.

Table 2.20: Race statistics for the CCD2 Dataset used in Chapter 2.

Attribute		Dataset Label	Subject Number	Left Images	Right Images
Race	**Caucasian**	White	33	340	360
	Non-Caucasian	Asian	23	150	120
		Black	3	10	30
		Other	17	100	90

Table 2.21: Gender statistics for the ND-GFI Dataset used in Chapter 2.

Attribute		Subject Number	Left Images	Right Images
Gender	Male	1500	750	750
	Female	1500	750	750

2.5.3 ND-GFI Dataset

The ND-GFI dataset is a publicly available dataset that was assembled by researchers at Notre Dame University. It contains 3000 NIR ocular images, 1500 of which are from male subjects and 1500 from female subjects. There are 750 right and 750 left images for each of the aforementioned categories. The dataset was first used in [84] but was discovered to contain multiple images from the same subjects ('an average of about six images per subject' [85]). The dataset was corrected and used again in [85] where it was stated to contain images from 1500 *unique* subjects, 750 males and 750 females. The images were captured with a LG 4000 sensor [85] and are labeled with the gender of the subject.

An additional ND-GFI validation dataset was also available (also collected by Notre Dame) containing 3 images per eye of 324 subjects for a total of 972 left and 972 right NIR ocular images [85].

2.5.4 BioCOP2008 Dataset

The images in the BioCOP2008 dataset were obtained using a near-infrared (NIR) sensor. Using a commercial off-the-shelf iris SDK, the center of the iris and its radius were automatically located. The iris was then centered horizontally, and the image was geometrically scaled such that the iris had a fixed radius of 120 pixels. The scaled image was then cropped around the repositioned iris
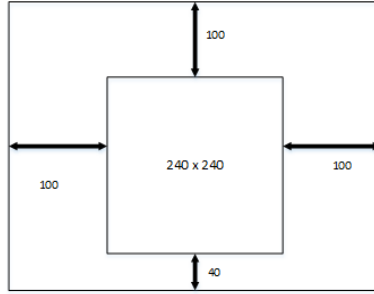
Figure 2.7: Geometrically adjusted ocular image for the BioCOP2008 dataset.

(a) Before   (b) After

Figure 2.8: Example of a geometrically adjusted image. Original image taken from [21].

Table 2.22: The subset of the BioCOP2008 iris dataset that was used in Chapter 2.

Attribute		Subject Number	Left Images	Right Images
Gender	Male	580	889	831
	Female	503	822	772

region so as to have a 40-pixel border below the iris and 100-pixel borders on the top and sides. The size of the scaled and cropped image was 440×380. See Figure 2.8. A total of 181 images, corresponding to about 5% of the entire dataset, were discarded during this step (for example, some images did not include the whole iris or could not be centered appropriately). The final dataset that was used consisted of 580 male subjects with 1720 images and 503 female subjects with 1594 images (please see Table 2.22 for a more complete breakdown). For each subject, images from both the left and right irides were included when available.

2.6 Experiments

2.6.1 BioCOP2009 Gender Results

Of the 1096 subjects contained in the BioCOP2009-1 dataset, 467 are labeled male and 629 are labeled female.[9] In order to assign an equal number of subjects to each class, 467 of the 629 available female subjects were randomly selected. The remaining 162 female subjects were not used for these experiments. 60% of the subjects were randomly chosen to be in the training set (280 subjects and their associated images) while the remaining 40% were placed in the test set (187 subjects and their associated images). This process of random selection was repeated 5 times, creating 5 different subject-disjoint sets for training and testing. An SVM classifier was trained on images from the training set. Images from all 3 sensors were pooled together during the training and testing process. Over the 5 random iterations for the left eye there were $10,727 \pm 3.4$ images used for training and $7,156 \pm 3.4$ images used for testing.[10] For the right eye there were $10,720 \pm 11.3$ images used for training and $7,159 \pm 11.3$ used for testing.[11] An SVM classifier was trained on each of the 5 training sets using the extracted BSIF features described in Section 2.4. The results of the experiments across all of the BSIF filter sizes are shown in Figure 2.9. The resulting confusion matrix is displayed in Table 2.25. An additional experiment was performed to determine the impact of training utilizing images that were flipped 180 degrees on the vertical axis and is shown in Table 2.23.

2.6.2 BioCOP2009-2 Race Results

The BioCOP2009-2 dataset used for the race prediction experiments in this subsection contains 247 subjects labeled with a variety of non-Caucasian classes (i.e., Asian, African American, American

---

[9]It should be noted that societal and personal interpretation of gender may consider more than a simple 'male' and 'female' label. For example, at the time of this publication, Facebook has 71 gender options.

[10]Some subjects may have more images than others.

[11]Some subjects may have more images than others.

Table 2.23: Gender prediction experiments performed on left eye images from the BiocCOP09-1 dataset. The results shown below are from an experiment that was performed to determine the impact of training not only the original images in the dataset, but also images that have been rotated 180 degrees around the vertical axis (i.e., flipped).

		Prediction Accuracy		
Random Iteration	**Training Set:**	**Original Only**	**Original and Flipped**	**Original and Flipped**
	Test Set:	**Original Only**	**Original Only**	**Original and Flipped**
1		85.93	84.62	84.66
2		85.63	84.18	84.18
3		85.10	84.21	84.33
4		84.65	82.88	82.88
5		85.89	84.61	84.89
Overall		85.44 ± 0.55	84.16 ± 0.78	84.12 ± 0.72
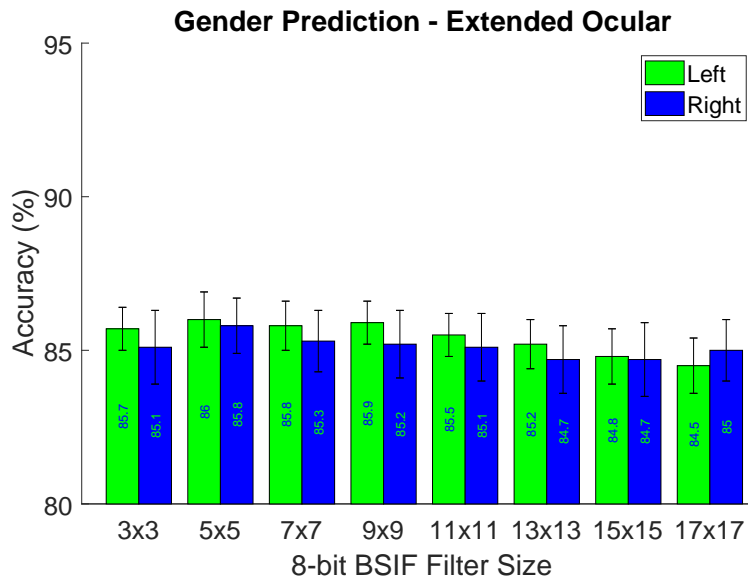
Figure 2.9: Gender prediction results using the extended ocular region (BioCOP2009-1 using 8-bit BSIF).

(a) Ocular Image   (b) Iris-Only Image   (c) Iris-Excluded Ocular Image
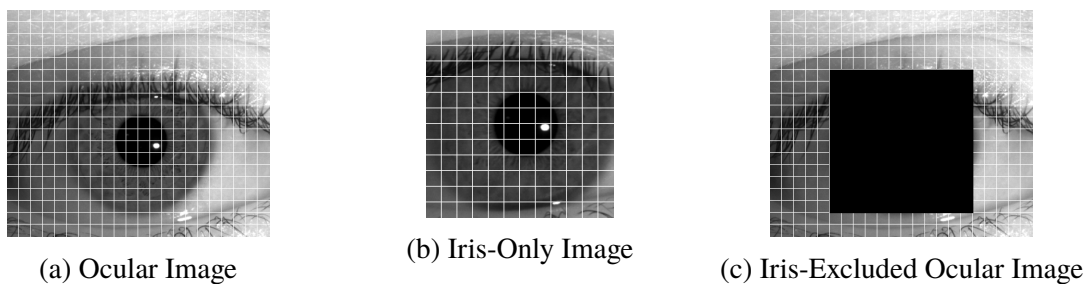
Figure 2.10: Tessellations applied to the three image regions. The images are from [21].

Table 2.24: Performance of the proposed gender prediction method on the BioCOP2009-1 dataset: BSIF 8-bit 9x9 filter size, LBP, LPQ.

Eye	Region	Gender		
		BSIF	LBP	LPQ
Left	Iris-Only	78.9 ± 1.0	78.9 ± 0.2	74.9 ± 1.0
	Iris-Excluded	82.2 ± 1.4	82.9 ± 0.9	81.8 ± 0.8
	Extended Ocular	85.9 ± 0.7	84.1 ± 0.5	82.4 ± 0.8
Right	Iris-Only	79.2 ± 0.8	79.8 ± 1.1	74.9 ± 1.2
	Iris-Excluded	82.1 ± 0.9	82.0 ± 0.6	80.6 ± 1.4
	Extended Ocular	85.2 ± 1.1	84.0 ± 0.7	81.4 ± 1.2

Table 2.25: Gender prediction confusion matrix for the extended ocular region (BioCOP2009-1 using 8-bit BSIF with a 9x9 filter).

	Left		Right	
	Predicted Female	Predicted Male	Predicted Female	Predicted Male
Actual Female	83.2% ± 1.8%	16.8% ± 1.8%	82.2% ± 2.2%	17.8% ± 2.2%
Actual Male	11.4% ± 1.7%	88.6% ± 1.7%	11.8% ± 1.8%	88.2% ± 1.8%

Indian, Hispanic, Middle Eastern, Asian Indian). In order to create an equal number of subjects in each of the classes, 247 of the 781 Caucasian subjects were randomly selected. The remaining Caucasian subjects were not used in the race prediction experiments. 60% of the subjects were randomly selected to be in the training set while the remaining 40% were selected for the test set, resulting in 148 subjects for training and 99 subjects for testing. This random selection was repeated 5 times resulting in 5 subject-disjoint training and testing sets. An SVM classifier was trained using the images from the 148 subjects selected for the training partition. Images from all 3 sensors were used during the training and testing stages. Over the 5 random iterations, there were 5656 ± 34 images in the training dataset and 3749 ± 34 images in the test set.[12]

The 8-bit BSIF was used in this chapter as a compromise between prediction accuracy and computational processing time. While 9-bit or 10-bit BSIF may provide slightly better results, the increased requirement of memory and processing time to perform each experiment was quite substantial given the large size of the BioCOP2009-2 dataset. An SVM classifier was trained on

---

[12]Some subjects may have more images than others.
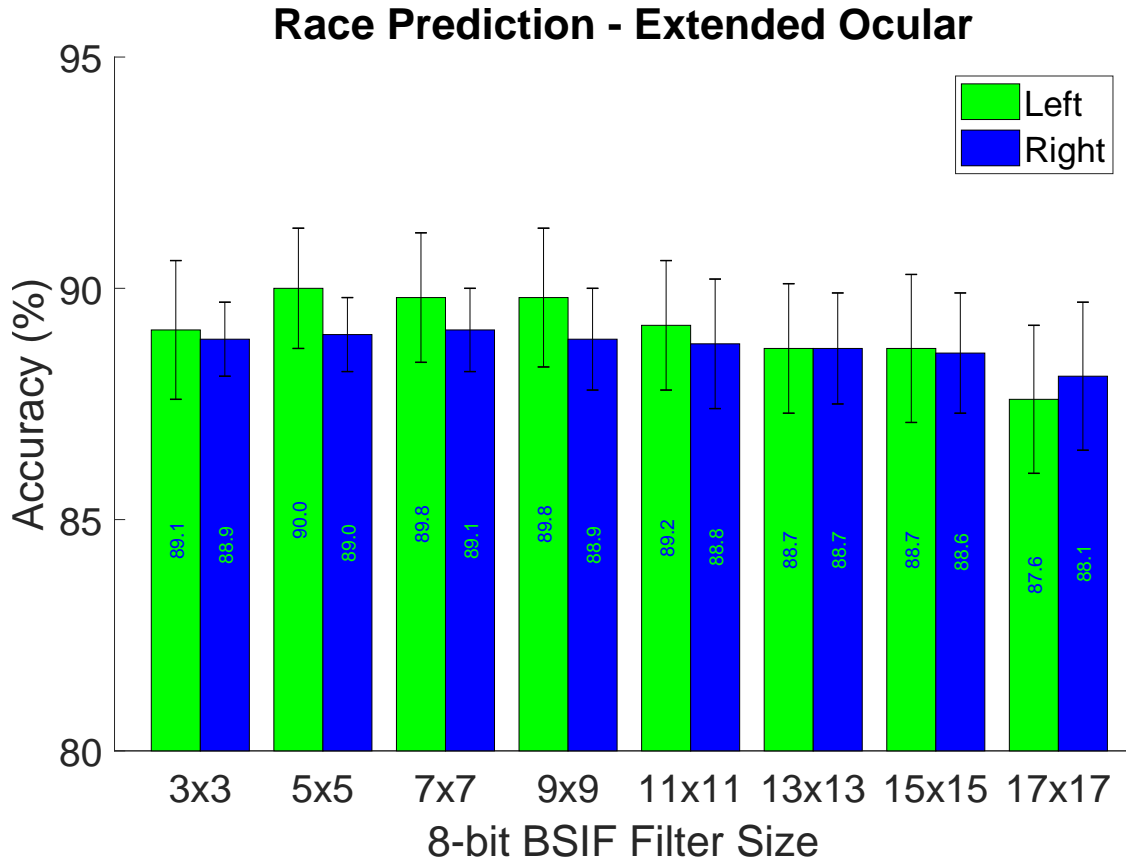
Figure 2.11: Race prediction results using the extended ocular region (BioCOP2009-2 using 8-bit BSIF).

Table 2.26: BioCOP2009-2 Race Texture Descriptor Comparison: BSIF 8-bit 9x9 filter size, LBP, LPQ.

		Race		
Eye	**Region**	**BSIF**	**LBP**	**LPQ**
	Iris-Only	88.9 ± 1.4	86.5 ± 1.5	86.9 ± 1.4
Left	Iris-Excluded	82.6 ± 1.5	88.0 ± 1.5	79.6 ± 0.9
	Extended Ocular	89.8 ± 1.5	88.4 ± 1.7	87.6 ± 1.3
	Iris-Only	88.6 ± 1.2	85.9 ± 0.8	87.1 ± 0.8
Right	Iris-Excluded	82.7 ± 0.6	88.0 ± 1.4	79.2 ± 0.6
	Extended Ocular	88.9 ± 1.1	87.1 ± 0.9	87.5 ± 0.8

each of the 5 training sets using the extracted BSIF features described in Section 2.4. The test data was classified using the respective SVM model. The resulting prediction accuracy using filter sizes in the range of 3×3 to 17×17 is shown in Figure 2.11. The resulting confusion matrix is displayed in Table 2.27.

Table 2.27: Race prediction confusion matrix for the extended ocular region (BioCOP2009-2 using 8-bit BSIF with a 9x9 filter).

	Left		Right	
	Predicted Caucasian	**Predicted Non-Caucasian**	**Predicted Caucasian**	**Predicted Non-Caucasian**
Actual Caucasian	91.7%1.3 ± %	8.3% ± 1.3%	90.5% ± 1.9%	9.5% ± 1.9%
Actual Non-Caucasian	12.1% ± 2.8%	87.9% ± 2.8%	12.8% ± 2.7%	87.2% ± 2.7%

Table 2.28: Gender prediction results using the Iris-Excluded and Iris-Only regions (BioCOP2009-1 BSIF 8bit-9x9 filter size).

Gender Prediction Accuracy (%)		
Eye	**Iris-Only Accuracy**	**Iris-Excluded Ocular Accuracy**
Left	78.9 ± 1.0	82.2 ± 1.3
Right	79.2 ± 0.8	82.1 ± 0.9

Table 2.29: Race prediction using the Iris-Excluded and Iris-Only regions (BioCOP2009-2 using 8-bit BSIF with a 9x9 filter).

Race		
Eye	**Iris-Only Accuracy**	**Iris-Excluded Ocular Accuracy**
Left	88.9 ± 1.4	82.6 ± 1.5
Right	88.6 ± 1.2	82.7 ± 0.6

2.6.3 Iris-excluded Ocular Region vs. Iris-Only Region

Previous work in this field [84, 85] has predominantly focused on the iris-only portion of the captured NIR ocular images. Bobeldyk and Ross [8] showed, for gender prediction using BSIF, that the ocular region provides greater sex prediction accuracy than the iris-only region. A separate feature vector was generated from each of the two regions: iris-excluded ocular and iris-only (see Figure 2.1). The results for race prediction are displayed in Table 2.29 and gender prediction in Table 2.28. For race, the iris-only region provides a *greater* prediction accuracy using BSIF than the iris-excluded ocular region, while the opposite is true for gender prediction.

| (a) Male | (b) Male | (c) Female | (d) Female |
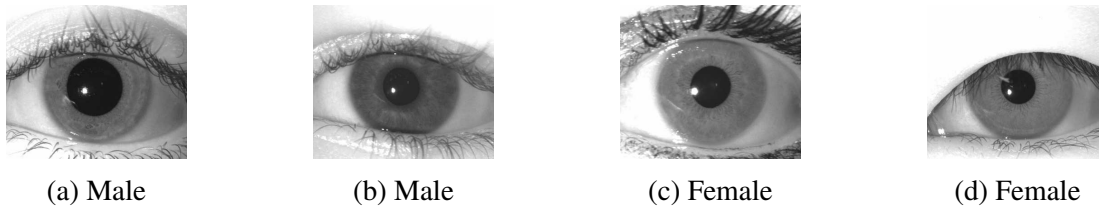
Figure 2.12: Misclassified images: (a) and (b) were classified as female, (c) and (d) were classified as male. The images are from [21].

2.6.4 Gender Cross Dataset Testing

In order to validate the proposed method and ensure generalizability of the algorithm to images originating from outside of the BioCOP2009-1 dataset, we chose to cross test on the following datasets: CCD1, CCD2, ND-GFI, and ND-GFI-validation. Each of these datasets were made available by the researchers at Notre Dame [21]. It was important to choose a dataset originating from a separate location than where the BioCOP2009-1 dataset was collected [13] in order to reduce the chances of the same identity being included in both of the datasets. The CCD1 and CCD2 datasets provide both gender and race labels for each of the images, while the ND-GFI and ND-GFI-validation datasets provides only gender. The CCD1 and CCD2 datasets contains images of subjects with contacts, without contacts and with cosmetic contacts. Only the images of subjects without contacts were used in the experiments.

The 5 trained SVM models that were generated using the BioCOP2009-1 dataset were used to classify images in each of the 4 selected datasets (CCD1, CCD2, ND-GFI, ND-GFI-validation). The results are shown in Table 2.30. The prediction accuracy for classification of images from CCD1 and CCD2 was about 10% less than those on the ND-GFI and ND-GFI-validation datasets. The authors believe this may be due to the increased number of images per subject in the cosmetic contact dataset. The images in the ND-GFI dataset, on the other hand, contain only 1 image per subject. Some images that were misclassified are shown in Figure 2.12.

---

[13]The BioCOP2009-1 dataset was collected at West Virginia University.

Table 2.30: Gender prediction results in a cross-dataset scenario where training and testing are done on different datasets (8-bit BSIF with a 9x9 filter).

Gender			
Training	**Testing**	**Eye**	**Prediction Accuracy**
BioCOP2009	CCD1	Left	75.3 ± 2.1
		Right	76.8 ± 2.9
	CCD2	Left	72.3 ± 4.0
		Right	77.8 ± 4.1
	ND-GFI	Left	84.4 ± 0.8
		Right	84.3 ± 0.5
	ND-GFI-Validation	Left	84.2 ± 1.2
		Right	82.6 ± 1.3

Table 2.31: Race cross dataset testing (8-bit BSIF with a 9x9 filter).

Race			
Training	**Testing**	**Eye**	**Prediction Accuracy (%)**
BioCOP2009	CCD1	Left	80.2 ± 1.3 [14]
		Right	90.3 ± 1.7
	CCD2	Left	87.3 ± 4.5
		Right	90.8 ± 1.6

2.6.5 Race Cross Dataset Testing

It is not uncommon for a method to perform well when training and testing are conducted using the same dataset. In order to demonstrate the generalizability of the proposed algorithm, we trained on the BioCOP2009-2 dataset and tested on the CCD1 and CCD2 datasets described earlier. The 5 trained SVM models that were generated using the BioCOP2009-2 dataset were used to classify the images in CCD1 and CCD2. It should be noted that subjects from the BioCOP2009-2 dataset were labeled as 'Caucasian' while those in the CCD1 and CCD2 datasets were labeled as 'White'. Both the CCD1 and CCD2 datasets contains images of people with contacts, without contacts and with cosmetic contacts. Only the images without contacts were used in our experiments. The results are shown in Table 2.31. Some images that were misclassified are shown in Figure 2.13.

---

[14] The lower prediction accuracy of the left eye could be attributed to the non symmetric composition of the subject pool between left and right eye images (of subjects that are not wearing contacts). If the contact lens images are also included, the prediction accuracy increases to 88.9% ± 1.2%).

| (a) Non-Caucasian | (b) Non-Caucasian | (c) Caucasian | (d) Caucasian |
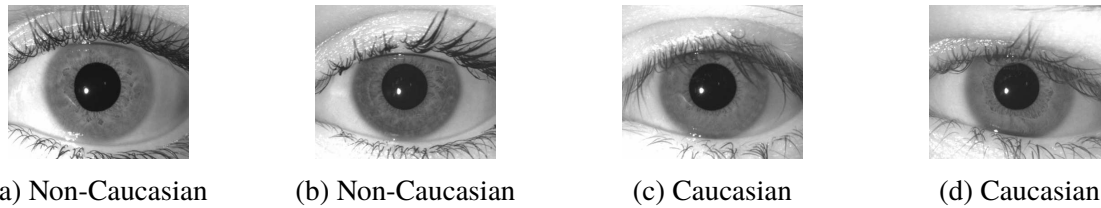
Figure 2.13: Misclassified images: (a) and (b) were classified as Caucasian, (c) and (d) were classified as Non-Caucasian. The images are from [21].

2.6.6 Impact of Race on Gender Prediction

In order to determine if predicting gender is a more challenging problem for either Caucasians or Non-Caucasians, 4 additional experiments were performed: (a) training and testing on Caucasian subjects; (b) training on Caucasian subjects and testing on Non-Caucasian subjects; (c) training and testing on Non-Caucasian subjects; (d) training on Non-Caucasian subjects and testing on Caucasian subjects.

Training and testing on only the Caucasian class results in a ~6% increase in prediction accuracy when compared to training and testing on only the Non-Caucasian class. The decrease in prediction accuracy for the Non-Caucasian class could be attributed to the multiple race labels that were assigned to the Non-Caucasian class (see Section 2.6.2). The results are shown in Table 2.32.

Training on either race class and cross testing on the other race class results in an ~80% prediction accuracy. It can be observed that there is a slight increase in prediction accuracy when training on the Non-Caucasian class and testing on the Caucasian class (~1-2%). The results are shown in Table 2.32.

2.6.6.1 Impact of Race on Gender Prediction - Additional Constraint

An additional constraint is imposed on the experiments from Section 2.6.6, where only subjects utilized to train the intra-race model are used to train the inter-race model (as opposed to using images from all of the Caucasian/Non-Caucasian subjects). Essentially, this will decrease the number of training images for the inter-race experiments and allow for a more equivalent comparison of the inter-race and intra-race prediction accuracies. The results for the left ocular images are shown in

Table 2.32: Gender prediction results for intra-race and inter-race training and testing (BioCOP2009-1 using 8-bit BSIF with a 9x9 filter).

	Gender		
Train On	**Test On**	**Eye**	**Prediction Accuracy**
Caucasian	Caucasian	Left	87.9 ± 1.3
		Right	87.2 ± 1.1
	Non-Caucasian	Left	77.5
		Right	78.5
Non-Caucasian	Non-Caucasian	Right	81.3 ± 2.5
		Left	81.2 ± 2.4
	Caucasian	Left	79.6
		Right	79.8

Table 2.33: Gender prediction results for inter-race training and testing (BioCOP2009-1 using 8-bit BSIF with a 9x9 filter). The results show that increasing the number of training subjects and images, increases the prediction accuracy.

	Gender			
Train On	**Number of Training Subjects**	**Test On**	**Eye**	**Prediction Accuracy**
Caucasian	Limited	Non-Caucasian	Left	76.2 ± 1.8
Caucasian	All	Non-Caucasian	Left	77.5
Non-Caucasian	Limited	Caucasian	Left	77.8 ± 0.6
Non-Caucasian	All	Caucasian	Left	79.6

Table 2.33.

It can be observed, that increasing the total number of subjects/images available for training, increases the prediction accuracy for both experiments (see Table 2.33). The outcome from this experiment is not unexpected as prediction models tend to benefit from a larger number of training images. It can also be observed, while probably not significant, that imposing this additional constraint would create a slightly larger gap in the prediction accuracies displayed in Table 2.32. In addition, given the limited size of the dataset and the inability to continue to increase the number of training subjects/images, the upper bound on race and gender prediction may still not be known.

Table 2.34: Race prediction results for intra and inter gender class training and testing (BioCOP2009-2 using 8-bit BSIF with a 9x9 filter).

Race			
Train On	**Test On**	**Eye**	**Prediction Accuracy**
Male	Male	Left	92.9 ± 2.1
		Right	92.2 ± 1.0
	Female	Left	78.6
		Right	78.7
Female	Female	Right	87.0 ± 1.8
		Left	88.9 ± 2.0
	Male	Left	88.3
		Right	88.3

2.6.7 Impact of Gender on Race Prediction

In order to determine if predicting race was a more challenging problem for either males or females, 4 additional experiments were conducted: (a) training and testing on Male subjects; (b) training and testing on Female subjects; (c) training on Male subjects and testing on Females; and (d) training on Females and testing on Males.

Training and testing on only male subjects results in a ~3-5% increase over training and testing on only female subjects.[15] There was a significant decrease in prediction accuracy when training on male subjects and testing on female subjects (~14%). There was no decrease in prediction accuracy when training on female subjects and testing on male subjects. The absence or presence of makeup in the female images may make it more difficult for the male-only trained model to predict race from the female images, but additional research should be performed to fully explore the difference in prediction accuracies. The results are summarized in Table 2.34.

---

[15]These prediction results agree with the findings of the earlier work from Lagree and Bowyer [46] who observed an increase in prediction accuracy when training and testing on male subjects, compared to female subjects.

Table 2.35: Eye color statistics by ethnicity and gender for the BioCOP 2009-1 dataset.

Eye Color	Caucasian		Non-Caucasian		Male		Female	
	Subjects	Images	Subjects	Images	Subjects	Images	Subjects	Images
Brown	267	10,330	228	8470	235	8,983	260	9,817
Blue	294	11,157	2	87	119	4,638	177	6,606
Green	137	5,251	6	226	46	1,806	97	3,671
Hazel	130	5,055	6	226	50	1,955	86	3,326
Gray	8	294	0	0	3	119	5	175
Other	0	0	18	734	14	543	4	191

Table 2.36: Impact of eye color on gender prediction (BioCOP2009-1 using 8-bit BSIF with a 9x9 filter).

Gender Prediction Accuracy (%)				
	Male		Female	
Eye Color	Left	Right	Left	Right
Brown	86.8 ± 3.4	87.0 ± 2.33	79.0 ± 3.4	81.0 ± 2.2
Blue	92.2 ± 1.1	84.6 ± 1.60	84.7 ± 3.9	85.2 ± 1.5
Green	93.0 ± 1.1	91.2 ± 2.15	84.7 ± 3.5	88.9 ± 2.4
Hazel	87.1 ± 2.5	85.0 ± 3.9	91.0 ± 2.6	85.7 ± 2.8

2.6.8 Impact of Eye Color on Race and Gender Prediction

The impact of eye color is also investigated on the prediction of race and gender from NIR ocular images.[16] The breakdown of eye color by ethnicity and gender for the BioCOP2009-1 dataset is listed in Table 2.35. The gender prediction accuracies categorized by eye color are shown in Table 2.36.

Gender Prediction: The results shown in Table 2.36 suggest that eye color does not have a significant impact on gender prediction. Males slightly outperform females regardless of eye color as seen in Table 2.36.

Race Prediction: The results shown in Table 2.37 suggest that eye color may have an impact on race prediction. Caucasian subjects with brown eyes have a lower prediction accuracy than Caucasian subjects with blue, green or hazel eye colors. Non-Caucasian subjects with brown eyes have a much greater prediction accuracy than non-Caucasian subjects with blue, green or hazel

---

[16]As in most iris data collection activities, eye color was self-declared by the subject and visually confirmed by the data collector.

Table 2.37: Impact of eye color on race prediction (BioCOP2009-2 using 8-bit BSIF with a 9x9 filter).

	Race Prediction Accuracy (%)			
	Caucasian		NonCaucasian	
Eye Color	**Left**	**Right**	**Left**	**Right**
Brown	79.1 ± 3.2	83.6 ± 2.9	90.4 ± 1.2	90.5 ± 1.5
Blue	98.5 ± 1.1	95.5 ± 0.6	0.0 ± 0.0	1.8 ± 2.2
Green	99.6 ± 0.3	90.2 ± 3.1	20.4 ± 11.7	6.0 ± 10.8
Hazel	90.1 ± 2.7	90.5 ± 3.6	64.1 ± 45.6	46.4 ± 34.2

eye colors. It should be noted that there only 14 non-Caucasian subjects without brown eyes (see Table 2.35).

2.6.9 Impact of Image Blur on Gender and Race Prediction

During the image acquisition process, ocular images may be captured out-of-focus. In order to determine the impact of out-of-focus images on both gender and race prediction, an additional experiment was performed. Out-of-focus images were simulated by 'blurring' the image.

The blurring effect was generated by applying a Gaussian filter to each image in the test partition with different sigma values (σ = 2, 4, 6, 8, 10). Images with varying levels of blur are displayed in Figure 2.14. Only the images in the test partition were blurred, while the images in the training partition were not blurred. The same subject disjoint experimental protocol used in the previous sections was followed (see Sections 2.6.1 and 2.6.2). In addition, applying a Gaussian filter to the images will help determine the impact of removing the high frequencies from the image. The experimental results are displayed in Table 2.38.

The prediction accuracy of race appears to decay quite rapidly in contrast with that of gender for blurred images (see Table 2.38). Given that the iris-only region provides a much greater prediction accuracy using BSIF than the iris-excluded region, it could be that the race information is encoded in the finer detail of the iris texture and as the rate of blur increases, the race information may begin to be obscured. This conclusion also leads us to believe that the discriminatory race information is encoded in the higher frequency band.
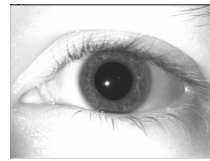
(a) Unmodified
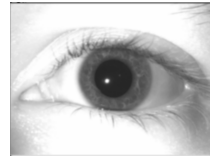
(b) $\sigma = 2$
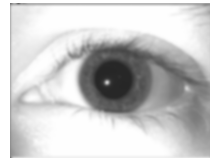
(c) $\sigma = 4$

(d) $\sigma = 6$

(e) $\sigma = 8$

(f) $\sigma = 10$

Figure 2.14: A sample ocular image that has been convolved with a Gaussian filter at different sigma values. The image in (a) is from [21].

2.6.10 Texture Descriptor Comparison

In order to select a suitable texture descriptor for experiments in this chapter, three were first considered: BSIF, LBP and LPQ. Each of the three texture descriptors are described in Section 4.3. Prediction accuracies were generated using the proposed methods from Section 2.6.1 and 2.6.2 for gender and race, respectively. The results of these experiments are shown in Tables 2.24 and 2.26.

Table 2.38: Gender and race prediction accuracy on blurred ocular images (BioCOP2009-1 for gender and BioCOP2009-2 for race using 8 bit BSIF with a 9x9 filter). Training is done on the original images in the train partition, while testing is done on the blurred images on the test partition.

Attribute	Attribute prediction accuracy from blurred images (%)					
	Unmodified	$\sigma = 2$	$\sigma = 4$	$\sigma = 6$	$\sigma = 8$	$\sigma = 10$
Gender	85.9 ± 0.7	83.1 ± 0.7	78.6 ± 1.1	75.3 ± 1.7	73.1 ± 2.6	70.8 ± 3.2
Race	89.8 ± 1.5	83.1 ± 1.5	64.8 ± 2.2	60.3 ± 2.5	58.4 ± 2.1	57.2 ± 2.0

BSIF was selected as the primary texture descriptor based on it's overall performance.

2.6.11 Convolutional Neural Network

Machine learning has been a dominant force in most recent computer vision and pattern recognition research. Given the success of utilizing a convolutional neural network (CNN) in several problems, a simple CNN is used in this chapter to determine its efficacy in predicting gender and race from NIR ocular images. The CNN used in this chapter is capable of predicting both gender and race with the architecture shown in Figure 2.15.

The initial CNN architecture that was considered was a down-scaled and modified version of Alexnet [42]. A number of different options were explored, including varying filter sizes and the number of layers. The final architecture that was used shown in Figure 2.15. Additional experiments on predicting gender and race utilizing a CNN can be found in Chapter 4.

In order to augment the number of images available to train the model, the training images from the BioCOP2009-1 and BioCOP2009-2 dataset were rotated by ±15 degrees with a step size of 3. All images were resized to 170×200 pixels. A subject disjoint training and testing protocol was used as discussed in section 2.6.1. The gender and race prediction accuracies using the proposed CNN are displayed in Table 2.40. While the CNN performs competitively, it was not able to outperform the texture descriptors discussed in the previous sections. An additional experiment was performed to determine the impact of data augmentation on gender prediction accuracy and is displayed in Table 2.39.

Table 2.39: Gender prediction experiments using a CNN trained from an augmented image dataset vs. a non-augmented image dataset. The experiment below shows a slight increase, about 1%, in prediction accuracy when the CNN model is trained with images from an augmented dataset. The augmented dataset contains images that were rotated ±15 degrees in steps of 3 using bilinear interpolation. An image size of 170×200 was used for the experiments (augmented and non-augmented). The time to train the model using the images from the augmented dataset took approximately 10 times as long. The prediction accuracy shown is from the first of five random subject partitions used for training and testing (as explained in Section 2.6.1).

Dataset	Gender Prediction Accuracy
Non Augmented	84.5%
Augmented	85.4%

Figure 2.15: CNN architecture for gender and race prediction from a NIR ocular image.

Table 2.40: Gender and race prediction accuracies utilizing the proposed CNN (BioCOP2009 dataset).

Attribute	Prediction Accuracy (%)	
	Left	Right
Gender	83.3 ± 1.7	82.4 ± 1.2
Race	85.9 ± 1.2	85.6 ± 1.0

(a) Ocular Image   (b) Iris-Only Image   (c) Iris-Excluded Ocular Image
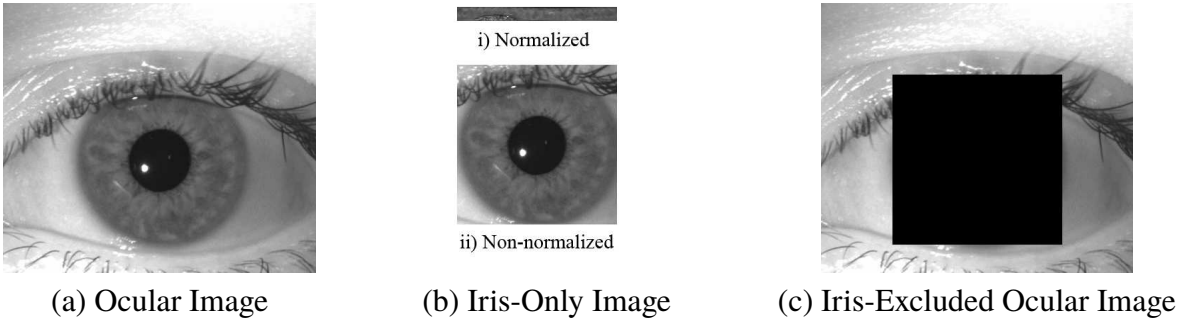
Figure 2.16: Four different regions of the ocular image considered for gender prediction. Original image taken from [21].

2.6.12 Normalized Iris Experiments

In addition to the 3 regions compared previously (see Figure 2.1), we conduct experiments to compare the sex prediction accuracy of the iris with the normalized iris region. The experiments in this section were conducted on the BioCOP2008 dataset (see Figure 2.5.4) and was a part of our earlier work in [8]. From each geometrically adjusted ocular image in the dataset, three different sub-images were extracted: iris-excluded ocular image, iris-only image, and normalized iris-only image. This resulted in the following four regions listed below:

Ocular Image: This is the entire scaled and cropped operational iris image (380×440). See Figure 2.16(a).

Iris-Only Image: The portion of the ocular image which encloses the entire iris region. The center of the image coincides with the center of the iris and the width of the image is twice the iris radius resulting in 240×240 images. No masking was performed to remove the eyelid or eyelash pixels. See Figure 2.16(b)(ii).

Normalized Iris-Only Image: The unwrapped iris-only image using Daugman's rubber sheet method [17]. The iris was sampled 20 times radially and 240 times angularly resulting in a 20×240 rectangular image. See Figure 2.16(a)(i).

Iris-Excluded Ocular Image: The ocular image with the iris-only region excluded. The portion of the image that was removed was zeroed out; essentially creating a black square in the middle of each of the images. See Figure 2.16(c).

50
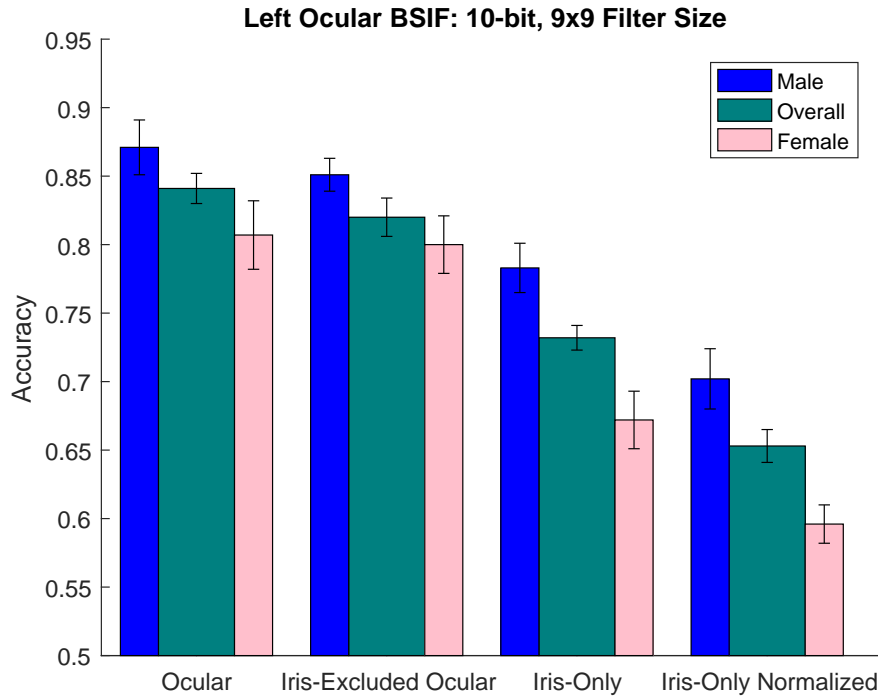
Figure 2.17: Results of the sex prediction accuracy for a particular combination of k and **n** on each of the four regions considered in our experiments using the BioCOP2008 dataset.

In order to capture both local and global spatial information, each image was tessellated into 20×20 blocks. Due to the small size of the normalized iris-only image, it was tesselated into 10×10 blocks. The BSIF operator was applied to the entire image and a histogram of the BSIF responses was computed for each block. Each histogram value was divided by the sum of the histogram values for that block, thereby normalizing it. The normalized histograms were concatenated together to form a feature vector that was input to a Support Vector Machine with a linear kernel.[17]

Each experiment was conducted using 60% of the subjects in the BioCOP2008 dataset for training and 40% for testing. This subject-disjoint partitioning exercise was done 5 times. Further, the impact of the number of filters (the bit length, **n**) and the size of each filter (k) on prediction accuracy was studied. Figure 2.17 report the accuracies corresponding to the four regions considered in this chapter. Results for the left and right eyes are shown separately in each figure. In each graph,

---

[17]For some combinations of k and **n**, the quadratic kernel outperformed the linear kernel. This has been noted in the legend of the performance graphs.

51
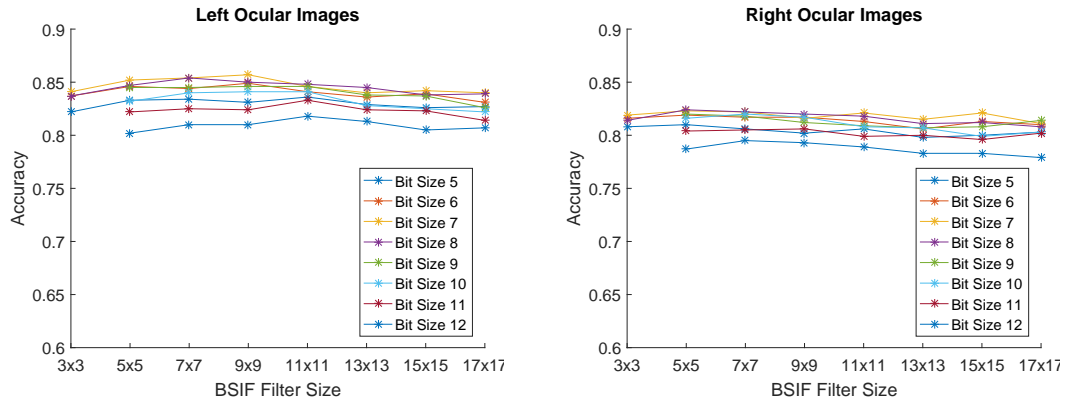
Figure 2.18: **Ocular image:** Sex prediction average accuracies for various BSIF bit lengths and filter sizes.
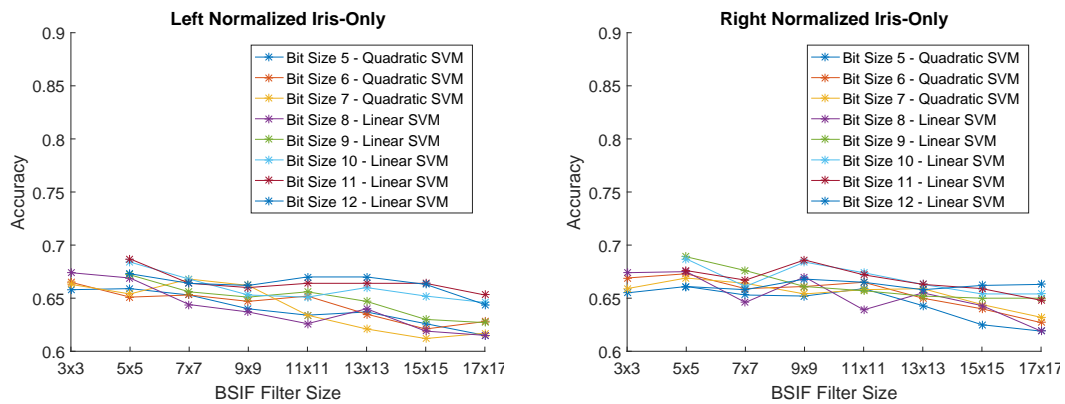
Figure 2.19: **Normalized iris-only image**: Sex prediction average accuracies for various BSIF bit lengths and filter sizes.
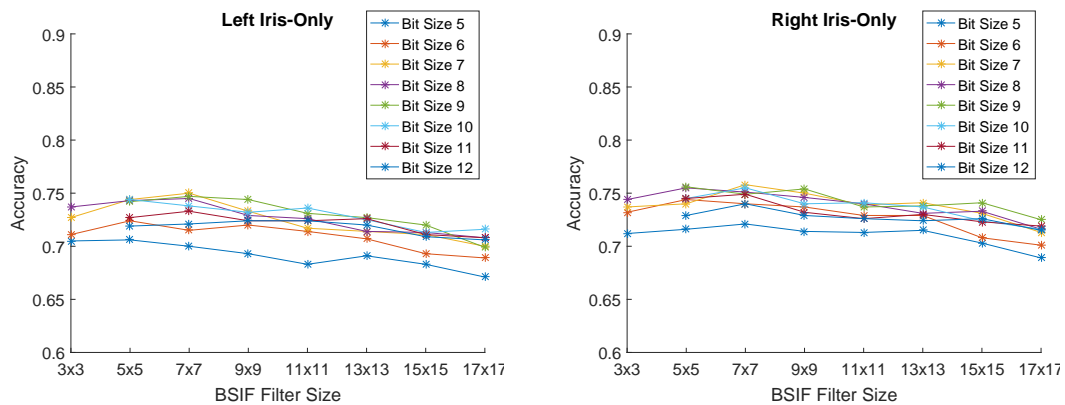
Figure 2.20: **Iris-only image:** Sex prediction average accuracies for various BSIF bit lengths and filter sizes.
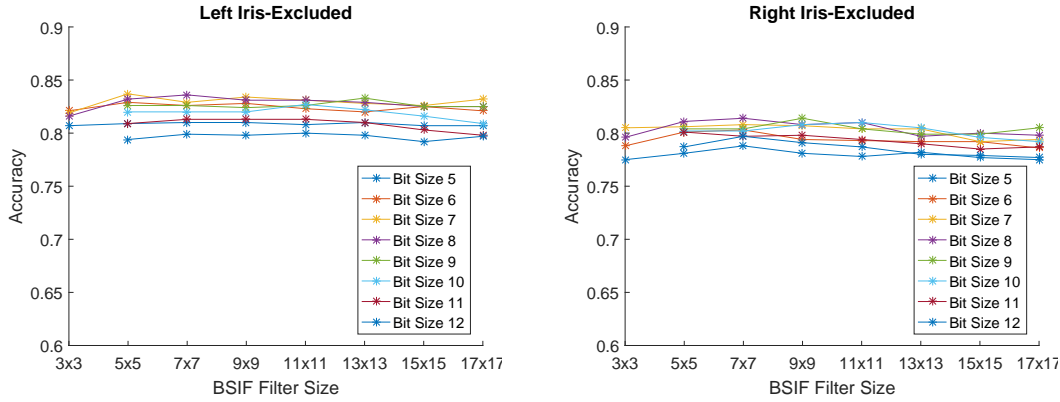
Figure 2.21: **Iris-excluded ocular image:** Sex prediction average accuracies for various BSIF bit lengths and filter sizes.

Table 2.41: Percentage of test images correctly/incorrectly classified by the iris-excluded ocular region **and** correctly/incorrectly classified by the iris-only region (Left eye, BSIF parameters: bit length = 10, filter size = 9×9).

| | | Iris-Only | |
		Correct	Incorrect
Iris-Excluded Ocular	Correct	$64.3 \pm 2\%$	$18.4 \pm 0.7\%$
	Incorrect	$8.7 \pm 1\%$	$8.6 \pm 0.4\%$

the average classification accuracy (over the 5 different trials) for different combinations of **n** and *k* is reported. While change in filter size does not seem to have a drastic impact on sex prediction for all 4 regions, change in bit length does have a somewhat discernible impact.

Figure 2.17 shows the male and female classification accuracies, along with the overall accuracy, for each of the four regions. The performance corresponds to the 10-bit BSIF operator with 9×9 filters. The ocular region and the iris-excluded ocular region exhibit the best performance while the normalized iris-only image exhibits the worst performance, with almost a 20% difference in performance over the ocular region. Further, the male classification accuracies are observed to be higher than the female classification accuracies. This could be partly attributed to the larger number of male subjects than female subjects in the dataset.

Table 2.41 shows the prediction relationship between the left iris-excluded ocular region and the left iris-only region. The values in this table are based on the 10-bit BSIF operator with 9×9 filters. The table indicates that there is a potential for fusing the outputs of these two regions which

Table 2.42: (Gender prediction accuracy from the normalized iris images on the CCD1 dataset with two different sampling sizes. 8-bit BSIF filter size 3×3).

Sample Size	Prediction accuracy
20×240	66.2 ± 2.2
30×360	70.3 ± 0.3

could possibly result in a higher overall prediction accuracy.

The non-normalized NIR iris image results in better sex prediction performance than the normalized iris region, thereby suggesting that the normalization process may be filtering out some useful information.

2.6.12.1 Normalized Iris Experiments - Increased Sample Size

An additional experiment was conducted to determine the impact of increasing the sampling used to generate the normalized iris image on gender prediction accuracy . Two sample sizes were considered: a) 20×240 using 10×10 tesselations and b) 30×360 using 15×15 tesselations. An increase in the tesselation size was used for the 30×360 size in order to generate feature vectors that were of equivalent length. The results are displayed in Table 2.42. The prediction accuracy does increase as the sample size also increases, which complements the findings of Section 2.6.12 by demonstrating the sampling process of generating the normalized iris image does appear to reduce the discriminatory information available. Future work should investigate increased incremental steps of the sample size to determine if there is a size that converges the non-normalized iris image.

2.7 Summary and Future Work

In this chapter, a number of experiments were performed to provide insight into the problem of predicting race and gender from NIR ocular images. Our broad findings are summarized below:

- **Texture Descriptors:** Gender and race prediction can be accomplished using simple texture descriptors. Both gender and race are predicted using the same feature vector (see Tables 2.24 and 2.26).

- **Generalizability:** The proposed algorithm is generalizable across multiple datasets and is, therefore, learning more than just artifacts from a single dataset. The generalizability applies for both gender and race prediction (see Tables 2.31 and 2.6.4).

- **Region Analysis:** The iris-excluded region provides greater prediction accuracy for gender than the iris-only region (see Table 2.28).

- **Non-normalized NIR iris image:** The non-normalized NIR iris image results in better sex prediction performance than the normalized iris region, thereby suggesting that the normalization process may be filtering out some useful information (see Table 2.17).

- **Left and Right:** There is no significant difference in performance between the left and right eye images for gender and race prediction (see Tables 2.24 and 2.31).

- **Cross-gender training:** For race prediction, training only on *male* images and testing on only *female* images results in a ~14% decrease in prediction accuracy than when training and testing on only male images. Training only on *female* images (also for race prediction) and testing on only *male* images shows no significant difference in prediction accuracy (see Table 2.34).

- **Impact of eye color on race and gender prediction:** For race prediction, Non-Caucasians with brown eyes displayed a higher prediction accuracy than Caucasians with Brown eyes. For gender prediction there was no observable impact based on eye color.

- **Impact of image blur on race and gender prediction:** The prediction accuracy for race degrades at a much faster rate than gender as the σ value of the Gaussian filter for blurring is increased.

Future work involving gender and race prediction from an NIR ocular image will add analysis of two additional covariates: 1) image scale and 2) salt and pepper noise. Image scale will be analyzed to determine the rate of prediction accuracy decrease as the scale of the image is decreased. The

second covariate, salt and pepper noise, will be induced in to each of the images to determine the rate of prediction accuracy drop off as the level of noise induced increases.

In addition to covariate analysis, small texture patches of the iris and surrounding ocular region will be analyzed. The goal of the analysis is to determine if certain texture patterns are prevalent in either gender, and if so, what do they look like? Also, is it possible that there is a correlation between the gender correlated texture patches and anatomical attributes of the iris or ocular region?

Increased sampling of the iris to generate the normalized iris image could also be further explored. This may determine if sampling greater than 30×360 would further increases the gender prediction accuracy.

Training on flipped right images as well as the left images (or vice versa) may be of interest to study. However, we have observed that CNN's appear to be laterality agnostic and so this experiment may not be relevant.

CHAPTER 3

PREDICTING EYE COLOR FROM NEAR INFRARED IRIS AND OCULAR IMAGES

The majority of this chapter has been published in the 11th IAPR International Conference on Biometrics (2018)

3.1 Introduction

Iris recognition systems typically acquire images of the iris in the near-infrared (NIR) spectrum rather than the visible spectrum. The use of NIR imaging facilitates the extraction of texture even from darker color irides (e.g., brown eyes). While NIR sensors reveal the textural details of the iris, the pigmentation and color details that are normally observed in the visible spectrum are subdued. In this chapter, we develop a method to predict the color of the iris from NIR images. In particular, we demonstrate that it is possible to distinguish between light-colored irides (blue, green, hazel) and dark-colored irides (brown) in the NIR spectrum by using the BSIF texture descriptor. Experiments on the BioCOP 2009 dataset containing over 43,000 iris images indicate that it is possible to distinguish between these two categories of eye color with an accuracy of ~90%. This suggests that the structure and texture of the iris as manifested in 2D NIR iris images divulges information about the pigmentation and color of the iris.

As stated before, the iris is typically imaged in the near-infrared (NIR) spectrum (as opposed to the visible spectrum which produces RGB images) for two primary reasons: (a) NIR illumination does not excite the pupil, thereby ensuring that the iris texture is not unduly deformed due to pupil dynamics during image acquisition [13]; and (b) the texture of dark-colored irides is better discerned in the NIR spectrum rather than the RGB color space, since NIR illumination tends to penetrate deeper into the multi-layered iris structure [11]. Therefore, NIR images capture the texture and morphology of the iris, but not the color of the iris. Sample images of the iris captured in both the NIR and the RGB color space can be seen in Figure 3.1.

It may seem implausible if not impossible to predict the 'eye color'[1] of an individual based on

---

[1]**Perceived** eye color is perhaps a more accurate term, as the color of an individual's eye can
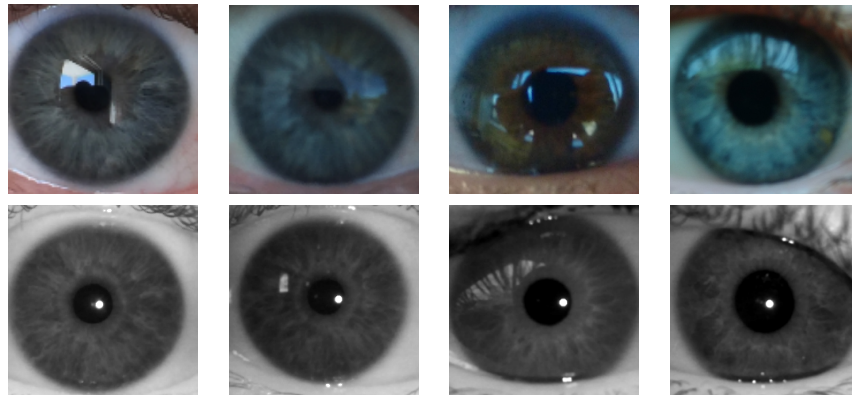
(a) Light color irides
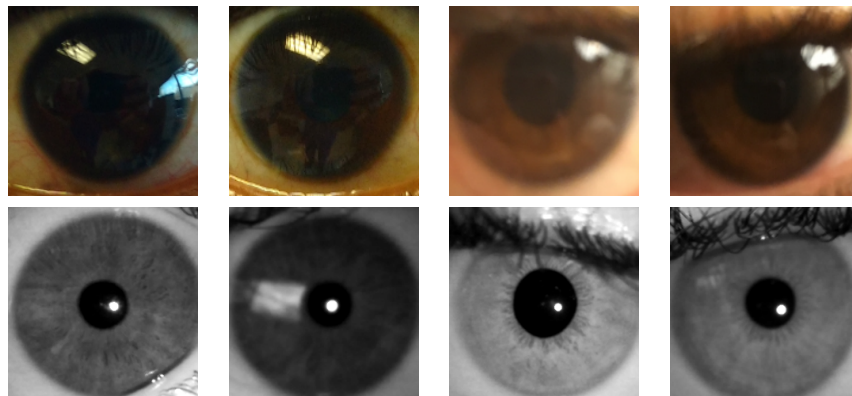
(b) Dark color irides

Figure 3.1: Examples of (a) light color irides, and (b) dark color irides. In each case, the top row shows images in the RGB color space and the bottom row shows the corresponding images in the NIR spectrum. The NIR images were taken with the Iritech IrisShield USB sensor while the RGB images were taken with a mobile camera. Notice that directly utilizing intensity information of the NIR images will not allow us to determine the pigmentation level of the iris.

NIR images. However, the texture and structure of the iris in the NIR spectrum can offer some cues about the pigmentation levels in the iris as described below.

3.2 Iris Pigmentation

There are 5 cell layers that make up the iris: the anterior border layer, the stroma, the sphincter muscle, the dilator muscle and the posterior pigment epithelium. Melanocytes, that are located in the anterior border layer and the stroma, produce melanin that is one of the determinators of

appear to vary due to external factors such as ambient light and iridescence. Further, multiple color shades may be evident within a single iris making it difficult to unambiguously assign a single color label to an iris.
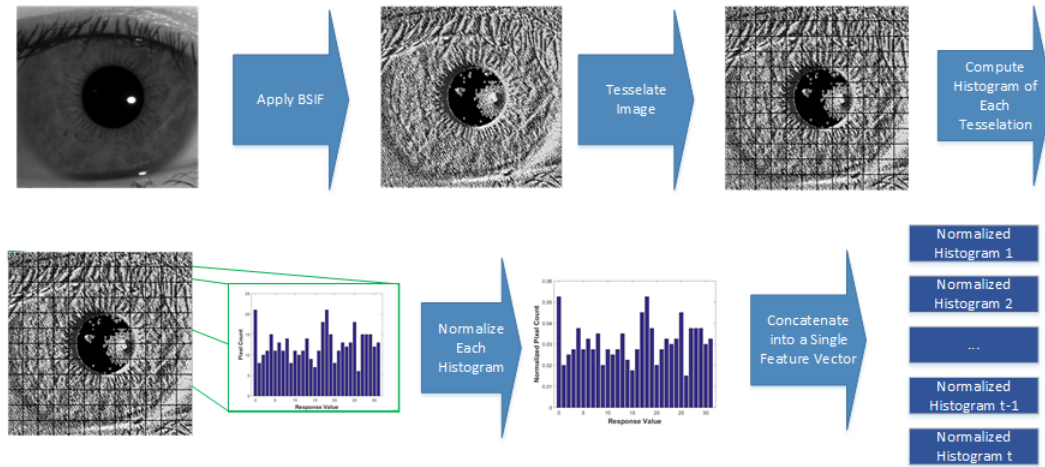
Figure 3.2: Generating the feature vector for eye color classification based on BSIF.

eye color. Darker color irides contain more melanin than lighter color irides [92]. The posterior pigment epithelium also contains melanin; however the amount of melanin in this layer is constant across different eyes, thereby not playing a significant role in the variation of eye color across the population [92]. The melanin in the anterior layer of darker color irides (i.e., brown) absorbs light as it passes through the cornea, reflecting back the brown color of the melanin. In lighter color irides (i.e., blue, green, hazel), the melanocytes contain little to no melanin. When the anterior layers contain little or no melanin, their *structure* will 'scatter the shorter blue wavelengths to the surface' [74]. This effect will makes the eye appear blue and is sometimes referred to as the 'Tyndale effect'.

Based on the foregoing discussion, we hypothesize that it may be possible to distinguish between dark color irides and light color irides in NIR images based on the structure of the iris. We assume that this structure of the iris is manifested in the textural nuances of the 2D NIR iris image. Therefore, we employ a texture descriptor to capture the structural information present in the iris. In particular, we employ a texture operator known as Binarized Statistical Image Features (BSIF) since it has been shown to outperform other descriptors in texture classification [40] as well as soft biometric prediction from NIR iris images [8]. The BSIF descriptor has also shown success in other iris biometric problems such as presentation attack detection [20, 62].

Benefits of this research: Predicting eye color from NIR iris images has several benefits and

possible applications: (a) Most legacy NIR iris datasets do not have information about eye color nor do they store the RGB image of the iris. Thus, predicting eye color from NIR images has both academic and practical utility; (b) Eye color can be used as an additional soft biometric cue for improving the performance of an iris recognition system via fusion or indexing [15]; (c) Eye color can also be used in cross-spectral matching scenarios, when comparing NIR iris images against RGB images [38]; (d) Assessing color and pigmentation level from NIR iris images would provide valuable insights into the correlation, if any, between iris pigmentation, iris color, iris texture and iris morphology; (e) Eye afflictions such as Pigment Dispersion Syndrome (PDS) can potentially be deduced from NIR iris images [66] if information about pigmentation levels can be ascertained; (f) Eye color can be used along with other soft biometric predictors to generate a semantic description of an individual (e.g., 'Asian middle-aged female with light colored eyes').

In this chapter, we will refer to eye images labeled[2] with the color 'brown' as category A, and eye images labeled as 'blue', 'green', 'hazel', or 'gray' as category B.[3] The rest of the paper is organized as follows: Section 3.3 discusses related work; Section 3.4 presents the two feature extraction methods used to predict eye color; Section 2.5.1.3 presents the dataset used; Section 3.5 presents the experiments and their results; Section 3.8 summarizes the findings of this work as well as discusses future work.

Table 3.1: Eye color prediction experiments utilizing 8 bit BSIF with a 3×3 filter on left eye images from the BioCOP09-4 dataset (see Section 2.5.1.4). Experimental results are shown below for the prediction of green, hazel and blue eye color. The anatomical literature typically categorizes green and hazel eye colors in the same category which could be an explanation for the poor prediction accuracy (when compared to the prediction accuracy for blue and brown).

Eye Color	Prediction Accuracy
Blue	85.9 ± 1.0
Green	69.1 ± 2.3
Hazel	71.7 ± 1.3

---

[2]The labels are typically self-declared by the subject during data collection and confirmed by the volunteer collecting the data.

[3]Additional experiments were performed in an attempt to predict 'blue', 'green' and 'hazel' eye color and those are shown in the Table 3.1.

3.3 Related Work

A careful review of the literature suggests that the topic of deducing eye color from *NIR images* has received limited attention. Dantcheva et al. [16] proposed an automatic system that detects eye color from standard facial images, but in the visible spectrum. They were interested in determining the viability of using eye color as a soft biometric for describing facial images. They also studied the impact of illumination, glasses, eye laterality as well as camera characteristics on assessing the eye color.

Howard and Etter [31] examined the impact of eye color on the identification accuracy of an NIR iris recognition system. Their work explored the impact of various attributes on match scores. They claimed that subjects with a certain ethnicity, gender and eye color had a higher false reject rate than other subjects in each of those categories (African American, female and black, respectively). They concluded that subject demographics and the impact of attributes on match scores can be used to develop subject-specific thresholds for recognition decisions. In relation to eye color, their work showed that persons with dark color irides exhibited a higher false rejection rate than persons with light color irides on a custom-built iris capture system based on a Goodrich/Sensors Unlimited 14-bit digital InGaAs camera.

However, none of the aforementioned work sought to predict eye color from NIR iris or ocular images.

3.4 Feature Extraction

As indicated earlier, we speculate that the pigmentation levels of the iris can be assessed from NIR images, thereby allowing us to determine the color of the eye. Such a hypothesis is based on our review of the eye anatomy literature which suggests that the melanin content (which is genetically determined) is correlated with the structure and texture of the iris [74, 92]. Thus, we use a histogram of filter responses to capture the local texture of the image, and an ordered enumeration of these histograms to capture the global structure of the iris (see Figure 3.2).

Two methods were used to generate the feature vector for eye color classification from NIR

images. The first method uses the texture descriptor BSIF. The second method uses the raw pixel intensity. The following two subsections detail the process used for each method.

3.4.1 Texture-based Method

Previous literature has demonstrated success in predicting both the gender and ethnicity of a subject using the texture of the iris and ocular region [8, 84]. The two texture descriptors that have performed particularly well in this context are Uniform Local Binary Patterns (LBP) and Binarized Statistical Image Features (BSIF). BSIF has been shown to outperform LBP in both the attribute prediction domain [8] and the texture classification domain [40]. Due to this reason, the BSIF descriptor was used in this work.

The BSIF descriptor was introduced by Kanala and Rahtu [40]. BSIF projects the input image into a subspace by convolving it with pre-generated filters. The pre-generated filters are created from 13 natural images supplied by the authors of [34]. 50,000 patches of size $k \times k$ are randomly sampled from the 13 natural images. Principal component analysis is applied, keeping only the top **n** components of size $k \times k$. Independent component analysis is performed on the top **n** components, generating **n** filters of size $k \times k$.

Each of the **n** filters is convolved with the input image and the ensuing response is binarized. The concatenated responses across the filters form a binary string that is converted into a decimal value (the BSIF response). For example, if the **n=5** binary responses are $\{1, 0, 0, 1, 1\}$, the resulting decimal value would be 19. Therefore, given **n** filters, the BSIF response will be in the interval $[0, 2^{\mathbf{n}} - 1]$.[4]

---

[4]While [40] states that the BSIF response is in the interval $[0, 2^{\mathbf{n}} - 1]$, the matlab code supplied by the authors utilizes a range of $[1, 2^{\mathbf{n}}]$.

Table 3.2: Number of subjects in each class used for training and testing.

Class	Subjects used for Training	Subjects used for Testing	Total number of subjects
Category A	297	198	495
Category B	297	286	583

(a) Captured Ocular Image   (b) Extracted and resized iris region
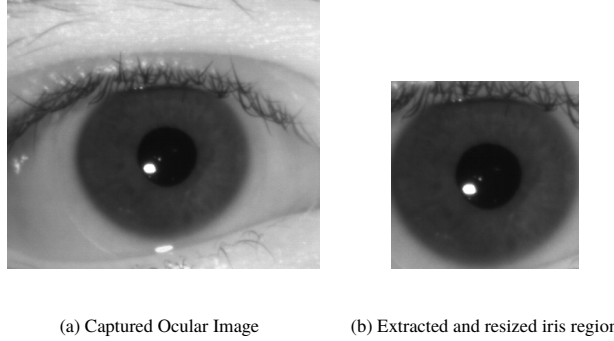
Figure 3.3: The iris region is extracted from the ocular image captured by the NIR sensor. Image taken from [21].

In order to provide consistent spatial information across images, the iris region in each image was cropped and resized to a 240×240 region (see Section 2.5.1.3 for details and Figure 3.3 for an example). The proposed texture-based method applies the BSIF operator to each NIR iris image. The filtered image is then tesselated into 20×20 pixel regions, for a total of 144 tessellations. This tessellation was performed in order to ensure that spatial order is encoded in the feature vector that is being created. A normalized histogram of length 2^{10} was generated for each of the 144 tessellations, and the histograms across all tessellations were concatenated into a single feature vector.

The parameters used for BSIF in our experiments were $\mathbf{n} = 10$ and $k = 7$. These parameter values were selected empirically based on [8]. Small-sized filters are more effective in capturing the local stochastic structure of the iris. The dimension of the texture-based feature vector was 147,456.

Table 3.3: Confusion matrix for the texture-based method (%).

	Left		Right	
	Predicted Category A	**Predicted Category B**	**Predicted Category A**	**Predicted Category B**
Actual Category A	88.7 ± 1.3	11.3 ± 1.3	88.9 ± 2.1	11.1 ± 2.1
Actual Category B	6.7 ± 0.8	93.3 ± 0.8	7.1 ± 0.9	92.9 ± 0.9

Table 3.4: Confusion matrix for the intensity-based method (%).

	Left		Right	
	Predicted Category A	**Predicted Category B**	**Predicted Category A**	**Predicted Category B**
Actual Category A	80.0 ± 1.3	20.0 ± 1.3	79.6 ± 1.6	20.4 ± 1.6
Actual Category B	18.2 ± 1.0	81.8 ± 1.0	17.4 ± 1.2	82.6 ± 1.2

3.4.2 Intensity-based Method

In order to generate a feature vector based on pixel intensity, each iris image was once again tesselated into 20×20 regions, resulting in a total of 144 tessellations. A histogram of the pixel intensities was generated for each of the 144 regions. The normalized histograms, each of length 256, were then concatenated into a single feature vector. The dimension of the intensity-based feature vector was 36,864.

The intensity-based method was considered in this work in order to determine if a dark color iris (or, respectively, a light color iris) in the RGB color space would manifest itself as dark (or light) in the NIR spectrum also. While Figure 3.1 provides visual evidence that this is *not* the case, it is worth confirming this in a rigorous manner.

3.5 Experiments

A subject-disjoint protocol was adopted to evaluate the proposed method. Therefore, subjects present in the training set did not have any of their images included in the test set, i.e., the subjects in the training and test sets were mutually exclusive. Further, both the training and test sets contained images from all 3 sensors.

60% of the subjects were randomly sampled to be used for training and the remaining 40% of the subjects were used for testing. This process was repeated 5 times in order to generate 5 separate partitions. Since some subjects have more images than others, the total number of training and

Table 3.5: Eye color prediction accuracy (%) using the feature vectors generated by the texture-based and intensity-based methods.

Eye	Texture-based	Intensity-based
Left	91.3 ± 0.8	81.1 ± 0.5
Right	91.3 ± 0.8	81.3 ± 0.6

Table 3.6: Eye color prediction accuracy (%) as a function of gender and ethnicity.

Method	Database Subset	Left Prediction Accuracy	Right Prediction Accuracy
Texture	Male	93.8 ± 1.0	93.7 ± 1.0
	Female	89.6 ± 1.0	89.5 ± 1.3
	Caucasian	90.3 ± 0.4	90.0 ± 0.6
	NonCaucasian	95.7 ± 2.0	96.4 ± 2.3
Intensity	Male	82.4 ± 0.6	82.8 ± 1.7
	Female	80.1 ± 0.9	80.3 ± 1.4
	Caucasian	79.4 ± 0.4	79.8 ± 0.7
	NonCaucasian	87.7 ± 1.3	87.4 ± 1.0

testing *images* varies across the five partitions. Since category B had a larger number of subjects than category A, category B training subjects were randomly sampled to equal the number of training subjects of category A. The additional subjects that were not used foBSIF, the prediction accuracy increases significantly over that of just using the feature extraction method based on the pixel intensity (9.1%) .

3.5.1   Texture-based Method

The feature vectors that were generated using the texture-based method (see subsection 3.4.1) were randomly partitioned by subject into 60% training and 40% testing as described above. The training feature vectors were used to create an SVM classifier (using a linear kernel). The SVM classifier was then used to predict the category to which each of the test feature vectors belonged to. This process was repeated for all 5 partitions, and the prediction accuracy results are shown in Table 3.5. The resulting confusion matrices for the left and right eye images are shown in Table 3.3.

3.5.2 Intensity-based method

The feature vectors that were generated from the intensity-based method (see Subsection 3.4.2) were randomly partitioned by subject into 60% training and 40% testing as described earlier. The training feature vectors were used to create an SVM classifier (using a linear kernel). The SVM classifier was then used to predict the category to which each test feature vector belonged to. The process was repeated 5 times and the resulting confusion matrices are shown in Table 3.4. The overall classification accuracy is shown in Table 3.5.

3.6 Eye Color Prediction Discussion

The prediction accuracy of the texture-based method outperforms that of the intensity-based method by 10% (see Table 3.5). This suggests that the intensity of NIR iris images cannot be solely used to predict eye color. Iris images from male subjects were found to have a slightly higher classification accuracy than those from female subjects for both the texture-based (~4%) and intensity-based (~2%) methods. There was very little difference in prediction accuracy between the left and right eye images (less than 1% in all cases).

3.7 Impact of Race and Gender on Eye Color

Further analysis of the experimental results, exhibit the prediction accuracy from non-Caucasian subjects is greater than those from Caucasian subjects; there was about a 6% difference using the texture-based method and about an 8% difference using the intensity-based method (see Table 3.5). We speculate this may be related to the higher number of non-Caucasian subjects in category A. Table 3.6 summarizes the results as a function of gender and ethnicity. Images from male subjects exhibit a slightly higher prediction accuracy than the images from female subjects. This slight discrepancy could possibly be investigated in future work.

3.8 Summary and Future Work

The focus of this chapter was on predicting eye color from NIR iris images. It is commonly assumed that eye color cannot be deduced from NIR iris images, since NIR illumination is not

well absorbed by melanin - the color inducing compound found in the iris. However, we show that texture and structure information evident in NIR images can be exploited to predict eye color. Two approaches were explored in this regard: a texture-based approach based on the BSIF texture descriptor, and an intensity-based approach based on raw pixel values. Experiments indicate that two categories of eye color can be distinguished with an accuracy of ~90% by the texture-based method. The intensity-based method performs substantially worse than the texture-based method, thereby suggesting that NIR pixel intensity does not accurately capture the notions of "dark color iris" and "light color iris" as observed in RGB color space.

Training and testing on each gender and race category exclusively as well as training on images of a single eye color and testing on images of a different eye color for gender and race prediction could be performed (similar to the experiments completed in Section 2.6.7 and Section 2.6.6). In addition to expanded gender and race analysis, the impact of image blur could be analyzed. Image scale and other common noises displayed in iris images could be two additional covariates to be analyzed. Image scale could be analyzed to determine the rate of prediction accuracy decrease as the scale of the image is reduced. Secondly, common noise in addition to image blur could be explored in order to determine types of noise that images in iris recognition systems typically experience. Each noise determined could be induced in to each of the images at varying levels to determine the rate of prediction accuracy drop off as the level of noise induction increases.

In addition to the covariate analysis, small texture patches of the iris region could be analyzed. The goal of the analysis is to determine if certain texture patterns are prevalent in either category A or category B subjects, and if so, what is their visual representation? Also, to determine if there is a correlation between the aforementioned eye color texture patches and a specific anatomical feature of the iris region (e.g., crypts, nevi)?

CHAPTER 4

IMPACT OF IMAGE SCALE ON ATTRIBUTE PREDICTION

The majority of this chapter was published in Winter Applications for Computer Vision Workshop 2019

4.1   Introduction

Most of the methods for extracting soft biometric attributes, have been evaluated on ocular images with reasonable resolution as characterized by their image size (see Table 4.1). In this chapter, we investigate the feasibility of extracting these attributes from low resolution images, i.e., ocular images that have been resized to a lower resolution (see Figure 4.1). Such a study has several potential benefits:

1. It would help establish the viability of predicting soft biometric attributes from poor quality input data; for example, iris images acquired at large standoff distances [55].

2. It would help in determining the degree of privacy offered by low resolution ocular images. In many applications, extracting soft biometric attributes without the subject's consent can be deemed to violate personal privacy [53]. Therefore, it is essential to ascertain the optimal image resolution that preserves privacy whilst permitting biometric functionality.

3. It can offer insight into the types of features being extracted by attribute prediction methods. Currently, there is a limited understanding of the precise cues that are being harnessed by automated methods for attribute prediction from iris or ocular images [43].

Table 4.1: The size of images in iris datasets that have been commonly used for research on attribute prediction.

Dataset	Image Resolution
ND Cosmetic Contact (e.g., [10])	480×640
ND-GFI (e.g., [85])	480×640
BioCOP2009 (e.g., [9])	480×640 and 480×480
UND_V (e.g., [85])	480×640
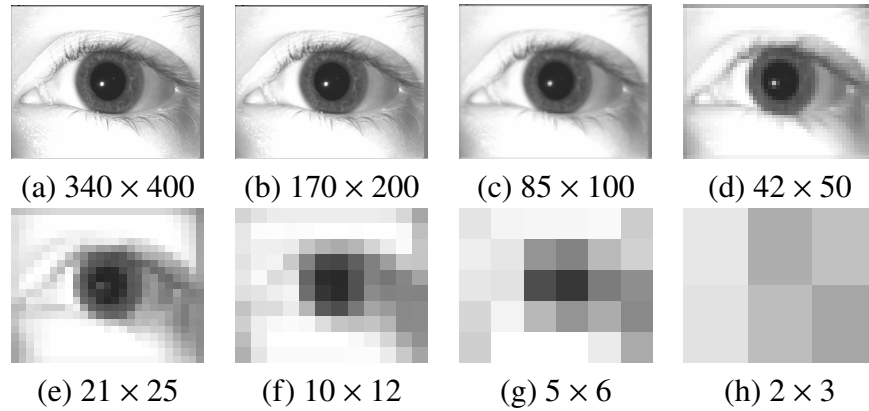
Figure 4.1: Multiple resolutions of a 480×640 source image. The source image was downsampled to the captioned image resolution, then displayed at a fixed size. The source image (not shown here) is from [21].

In this chapter, we will present the results of an experiment that will help assess the impact of image resolution on attribute prediction in the context of ocular images that are used for iris recognition. The prediction of three binary attributes, viz., gender, race and eye color, will be observed as a function of image resolution (see Figure 4.1). Two attribute prediction methods will be used in this regard: the first based on the BSIF (Binarized Statistical Image Features) texture descriptor and the second based on a Convolutional Neural Network (CNN). These attribute prediction methods were selected due to their observed success in recent literature [8, 85, 81, 9]. The rest of the chapter will briefly discuss related work (Section 4.2); introduce the feature extraction and classification methods (Section 4.3); present the datasets used for the experiments (Section 4.4); discuss the experiments conducted along with their results (Section 4.5); and conclude with a discussion of the findings of this chapter and future work (Section 4.6).

4.2 Related Work

We are not aware of any existing work in the iris and ocular biometrics literature that investigates the impact of image resolution on attribute prediction. There is, however, related work involving low quality or blurred iris images with a focus on increasing biometric recognition accuracy.

There are several techniques that involve taking a low resolution image and transforming it into a higher resolution one (see [57]). Fahmy [24] reconstructed a high resolution iris image from an

Table 4.2: BSIF-based method. The size of each BSIF filter, as well as the size and number of tessellations used, are indicated for each image resolution.

Image Size	Tessellation Size	Number of Tessellations	BSIF Filter Size
340×400	20×20	17×20	9×9
170×200	10×10	17×20	9×9
85×100	10×10	9×10	9×9
42×50	5×5	9×10	9×9
21×25	5×5	5×5	9×9
10×12	Whole Image	Whole Image	3×3
5×6	Whole Image	Whole Image	3×3
2×3	Whole Image	Whole Image	3×3

RGB video of low resolution images; however, there were no experiments performed to determine the impact of the method on either iris recognition or attribute prediction. Barnard et al. [6] used a multi-lens imaging system to compute a high resolution image from an ensemble of low-resolution images for the purpose of iris recognition. Huang et al. [33] enhanced the resolution of low resolution/blurry images by learning prior information from the different frequency bands. Their method increased the recognition rate of an iris recognition system. Instead of attempting to super resolve the image for visual purposes, Nguyen et al. [56] proposed to enhance the *features* used for recognition. Jillela et al. [39] used the principal components transform to perform image-level fusion of low-resolution iris images to improve recognition accuracy.

In contrast to existing work, our goal is to determine if there is sufficient information in low-resolution ocular images to permit the extraction of soft biometric attributes.

4.3   Feature Extraction

Two methods were employed to analyze the effect of image resolution on attribute prediction from NIR ocular images. The first method uses the hand-crafted feature descriptor BSIF to generate a feature vector that is input to a trained linear Support Vector Machine (SVM) classifier. The second method uses a simple two-layer CNN that serves the dual function of feature extraction and classification.

4.3.1 BSIF texture descriptor

BSIF was selected as the texture descriptor of choice based on the success it has shown in the iris attribute prediction domain [8, 10]. Using a commercial SDK, the center and radius of the iris were located in each image. The images were then cropped and resized to a fixed size (similar to [10]). Each image was convolved with the 8 pre-generated filters provided in [40], using 9×9 filters for the larger images and 3×3 filters for the smaller images. In order to preserve spatial information, each of the filtered images was tessellated into smaller regions. A 256-dimensional histogram was extracted from each region and the regional histograms were concatenated into a single feature vector. The number and size of tessellations used for each image resolution are shown in Table 4.2.

4.3.2 Convolutional Neural Network

A simple two-layer convolutional neural network (CNN) was used in our analysis, and the size and number of filters in each convolutional layer are shown in Table 4.3. Three and four-layer CNN models were also evaluated, with different number of filters and filter sizes, but the CNN presented in this chapter resulted in the best prediction accuracy. The input layer was modified to match the input image size. The architecture was kept consistent as much as possible across different image resolutions, except when the filter size exceeded that of the input image. A reduction in filter size was required for the smaller input image sizes (10×12, 5×6 and 2×3). The filter size was reduced to 3×3 in both convolutional layers for the 10×12 and 5×6 image sizes. For the 2×3 image size, the filter size was reduced to 2×2. In addition, the max pooling layer was removed from the first convolutional layer in this case. For the special case of the 1×1 image (a single pixel), no convolutional layers were used, see Section 4.5.6 for details.

4.4 Datasets

Three different datasets were used for the experiments, the BioCOP2009-4, BioCOP2009-5 and the ND Cosmetic Contact dataset [21]. The BioCOP2009-4 dataset, described in Section 2.5.1.4, was used for all of the intra-dataset experiments in this chapter, with the exception of the intra-dataset

Table 4.3: The CNN architecture used for each of the 8 input image resolutions.

| Image Size | Convolutional Layer One | | Max Pool Layer | Convolutional Layer Two | |
	Filter Size	Number of Filters		Filter Size	Number of Filters
340×400	5×5	48	Yes	6×6	128
170×200	5×5	48	Yes	6×6	128
85×100	5×5	48	Yes	6×6	128
42×50	5×5	48	Yes	6×6	128
21×25	5×5	48	Yes	6×6	128
10×12	5×5	48	No	3×3	128
5×6	3×3	48	No	3×3	128
2×3	2×2	48	No	2×2	128

race prediction experiments which used the BioCOP2009-5 dataset described in Section 2.5.1.5. The ND Cosmetic Conctact dataset was used to evaluate the generalizability of the results across datasets. The ND Cosmetic Contact dataset used in this chapter is described in Section 4.4.1.

4.4.1 Cosmetic Contact Dataset

In order to perform cross-dataset testing, we used the Cosmetic Contact Lens dataset assembled by researchers at the University of Notre Dame [21]. The Cosmetic Contact Lens dataset was selected due to its availability to the research community and contains images labeled with two of the three attributes that are explored in this chapter (race and gender). The dataset contains images collected by 2 separate sensors, the LG4000 and the AD100. For the LG4000 sensor, 3000 images were collected for training and 1200 images were collected for testing. For the AD100 sensor, 600 images were collected for training and 300 images were collected for testing. In this preliminary work, we only used the 3000 *training* images from the LG4000 sensor. No experiments were performed on the other images. The same geometric alignment process that was used for the BioCOP2009 images (see Figure 2.6) was applied to the Cosmetic Contact images. No images were discarded during this processing step. There are 1550 left ocular images: 550 images where the subject is wearing a cosmetic contact lens, 500 with contacts and 500 without contacts. There are 1450 right ocular images: 450 of the images where the subject is wearing a cosmetic contact lens, 500 with contacts and 500 without contacts.

4.5   Experiments

In order to determine the impact of image resolution on attribute prediction, two feature extraction methods were adopted: the first utilizing the handcrafted feature descriptor BSIF, the second method utilizing a convolutional neural network. The prediction accuracy of each of the three attributes, viz., gender, race and eye color, were determined at varying image resolutions. The largest image size used is the original cropped and resized image of 340×400 pixels. The bicubic interpolation algorithm was used to repeatedly scale down the image size by a factor of 4. The following image sizes were subsequently analyzed: $340 \times 400, 170 \times 200, 85 \times 100, 42 \times 50, 21 \times 25, 10 \times 12, 5 \times 6$ and 2×3. A special case scenario was also investigated where attribute prediction from just a single pixel was performed in Section 4.5.6.

Gender[1] was considered as a two class problem: 'male' or 'female'. Race was considered to be a two class problem: 'Caucasian' or 'Non-Caucasian'. Eye color was also considered as a two class problem, labeled 'category A' or 'category B'; see Section 4.4 for additional details.

A subject disjoint protocol was adopted, i.e., subjects in the training and test sets were mutually exclusive. For each attribute, 60% of subjects were used for training and 40% were used for testing. The number of subjects in the training partition for each class was balanced by randomly selecting subjects from the larger training class to equal that of the smaller training class. This process was repeated 5 times to generate 5 random partitions of the dataset for each of the three attribute prediction problems.

4.5.1   Simple Prototype-based Method

Is it possible there is a class prototype that can be used to exemplify each class of the binary attributes? In order to make this determination, the average image, for each attribute class, was calculated. The training images in each partition were used to compute the mean image and the

---

[1]It should be noted that societal and personal interpretation of gender may consider more than a simple 'male' and 'female' label. For example, at the time of this chapter's publication, Facebook has 71 gender options.
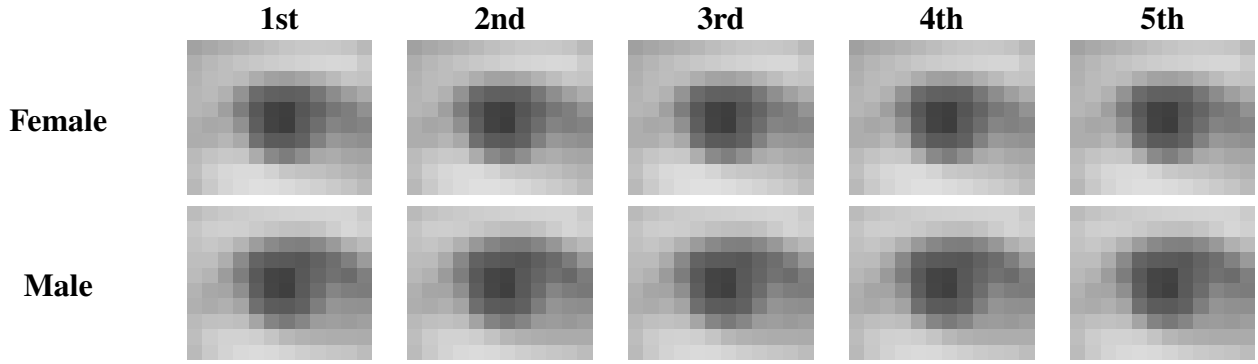
Figure 4.2: The 10×12 mean images from each of the 5 training partitions of the BioCOP2009 dataset used for the prototype-based classification method.

(a) Left

(b) Right
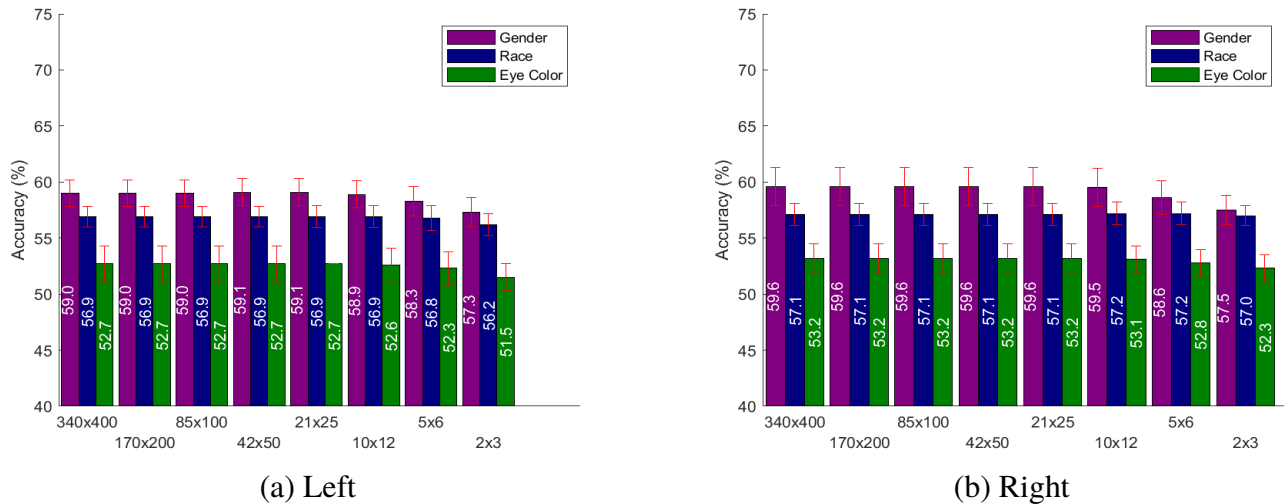
Figure 4.3: The naive prototype-based method. Attribute prediction accuracy (in %) at different image resolutions. The prediction accuracies are expectedly very low.

process was repeated for each of the 5 training partitions. The 10×12 mean images for the gender attribute are displayed in Figure 4.2. The attribute of a test image was then predicted by computing its Euclidean distance with the prototype image in each class, and assigning it to that class with the lowest distance. This method failed to demonstrate a prediction accuracy greater than 60% for *any* of the image sizes (see Figure 4.3). The conclusion that can be drawn from this experimental result is that raw pixel intensities alone do not contain sufficient discriminatory information to predict any of the 3 attributes.

Given the inability of the naive prototype-based method to perform attribute prediction, we next
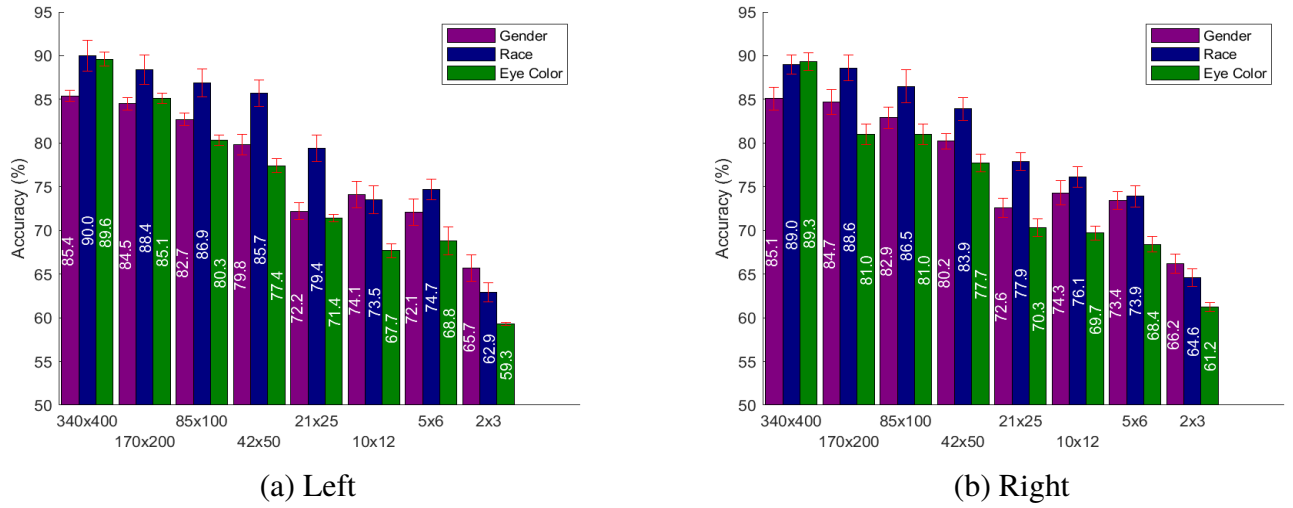
Figure 4.4: BSIF-based method. Attribute prediction accuracy (in %) at different image resolutions.

investigate if more sophisticated image representation and classification models can be appropriated for the study.

4.5.2 BSIF-based method

A feature vector was generated for each image using the method described in Section 4.3.1. An SVM classifier was generated using the feature vectors from the training partition for each of the three attributes. The test images were then classified using the SVM classifiers. This process was repeated for each of the 5 random partitions of the dataset. The resulting prediction accuracy of images from the test partition is displayed in Figure 4.4 and Table 4.4. The prediction accuracy continually decreases as the image resolution also decreases. For the left ocular images, at the largest image resolution of 340×400, the prediction accuracies for gender, race and eye color were $85.4 \pm 0.6\%$, $90.0 \pm 1.8\%$ and $89.6 \pm 0.8\%$, respectively. The prediction accuracies gradually decline to $72.1. \pm 1.5\%$, $74.7 \pm 1.2\%$ and $68.8 \pm 1.6\%$ for gender, race and eye color, respectively, at 5×6 image resolution. For the right ocular images, at the largest image resolution of 340×400, the prediction accuracies for gender, race and eye color were $85.1 \pm 1.3\%$, $89.0 \pm 1.1\%$ and $89.3 \pm 1.0\%$, respectively. The prediction accuracies gradually decline to $73.4 \pm 1.1\%$, $73.9 \pm 1.2\%$ and $68.4 \pm 0.9\%$ for gender, race and eye color, respectively, at 5×6 image resolution.

75

Table 4.4: BSIF-based method. Attribute prediction accuracy (in %) at different image resolutions.

Size	Gender		Race		Eye Color	
	Left	**Right**	**Left**	**Right**	**Left**	**Right**
340×400	85.4 ± 0.6	85.1 ± 1.3	90.0 ± 1.8	89.0 ± 1.1	89.6 ± 0.8	89.3 ± 1.0
170×200	84.5 ± 0.7	84.7 ± 1.4	88.4 ± 1.7	88.6 ± 1.5	85.1 ± 0.6	81.0 ± 1.2
85×100	82.7 ± 0.7	82.9 ± 1.2	86.9 ± 1.6	86.5 ± 1.9	80.3 ± 0.6	81.0 ± 1.2
42×50	79.8 ± 1.2	80.2 ± 0.9	85.7 ± 1.5	83.9 ± 1.3	77.4 ± 0.8	77.7 ± 1.0
21×25	72.2 ± 1.0	72.6 ± 1.1	79.4 ± 1.5	77.9 ± 1.0	71.4 ± 0.4	70.3 ± 1.0
10×12	74.1 ± 1.5	74.3 ± 1.4	73.5 ± 1.6	76.1 ± 1.2	67.7 ± 0.8	69.7 ± 0.8
5×6	72.1 ± 1.5	73.4 ± 1.0	74.7 ± 1.2	73.9 ± 1.2	68.8 ± 1.6	68.4 ± 0.9
2×3	65.7 ± 1.5	66.2 ± 1.1	62.9 ± 1.1	64.6 ± 1.0	59.3 ± 0.2	61.2 ± 0.5
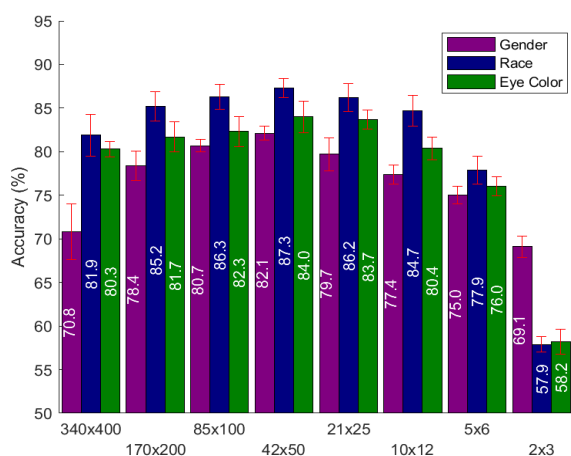
(a) Left
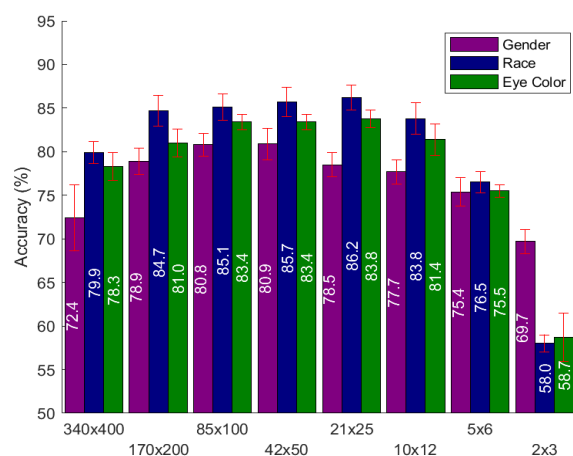
(b) Right

Figure 4.5: CNN-based method. Attribute prediction accuracy (in %) at different image resolutions.

4.5.3 CNN-based method

The first layer of the CNN is the image input layer followed by a convolutional layer of 48 feature channels with 5×5 filters. A ReLU layer receives the output from the convolutional layer and feeds it forward to a 3×3 max pooling layer with a stride of 2. The output is then forwarded to the second convolutional layer utilizing 128 feature channels with 6×6 filters. This output is then forwarded to a ReLU layer followed by a single fully connected layer, a softmax layer and finally a classification layer.

In order to be able to compare the prediction accuracies across the various image sizes, the architecture of the CNN was kept the same to the extent possible. The architectures for the smaller

Table 4.5: CNN-based method. Attribute prediction accuracy (in %) at different image resolutions.

Size	Gender		Race		Eye Color	
	Left	**Right**	**Left**	**Right**	**Left**	**Right**
340×400	70.8 ± 3.2	72.4 ± 3.8	81.9 ± 2.4	79.9 ± 1.3	80.3 ± 0.9	78.3 ± 1.6
170×200	78.4 ± 1.7	78.9 ± 1.5	85.2 ± 1.7	84.7 ± 1.8	81.7 ± 1.7	81.0 ± 1.6
85×100	80.7 ± 0.7	80.8 ± 1.3	86.3 ± 1.4	85.1 ± 1.5	82.3 ± 1.7	83.4 ± 0.9
42×50	82.1 ± 0.8	80.9 ± 1.8	87.3 ± 1.1	85.7 ± 1.7	84.0 ± 1.8	83.4 ± 0.9
21×25	79.7 ± 1.9	78.5 ± 1.4	86.2 ± 1.6	86.2 ± 1.4	83.7 ± 1.1	83.8 ± 1.0
10×12	77.4 ± 1.1	77.7 ± 1.4	84.7 ± 1.8	83.8 ± 1.8	80.4 ± 1.3	81.4 ± 1.8
5×6	75.0 ± 1.0	75.4 ± 1.6	77.9 ± 1.6	76.5 ± 1.2	76.0 ± 1.1	75.5 ± 0.7
2×3	69.1 ± 1.2	69.7 ± 1.4	57.9 ± 0.9	58.0 ± 1.0	58.2 ± 1.4	58.7 ± 2.8

image sizes (2×3, 5×6, 10×12) use a 3×3 filter size for both convolutional layers since the 5×5 and 6×6 filter sizes exceed the dimensions of these images.

The prediction accuracies for the three attributes were computed for all 5 test partitions. The results are displayed in Figure 4.5 and Table 4.5.

The performance, perhaps surprisingly, starts to increase at first as the image resolution decreases. For the left ocular images, the prediction accuracy for gender, race and eye color increases by 9.9%, 4.4% and 2%, respectively, when the image resolution is changed from 340×400 to 85×100. For the left ocular images, the prediction accuracy decreases for gender, race and eye color by only 3.6%, 2.3% and 4.7%, respectively, when the image resolution changes from 85×100 to 5×6. The prediction accuracies for gender, race and eye color for 5×6 images were 75 ± 1.0%, 77.9 ± 1.6%, 76 ± 1.0%, respectively. A similar phenomena occurs for the right ocular images (see Figure 4.5 and Table 4.5). In Section 4.5.5, we will further increase the performance of low resolution images by judiciously modifying the CNN.

One possible explanation for the reasonable performance of 5×6 images is, as the input to the network gets smaller, the number of parameters to be learned decreases; therefore, we technically have 'more data' for the smaller networks. If there were more training data available for the larger resolutions, the corresponding networks may have resulted in even better performance.

4.5.4 CNN-based method Cross-Dataset Testing

Given the superior performance of the CNN-based method compared to the BSIF-based method, we were interested in determining if the former would generalize across datasets. In order to test its generalizability, images from the Cosmetic Contact dataset were classified using a model trained only on images from the BioCOP2009-4 dataset for gender and the BioCOP2009-5 dataset for race. It must be noted that the label 'Caucasian' was used in the BioCOP 2009-5 dataset and the label 'White' was used in the Cosmetic Contact dataset. We assume these two labels to be equivalent. Given that the images from the Cosmetic Contact dataset had not been cropped and resized (see Section 4.3), the non-cropped and resized images from the respective BioCOP2009 datasets were used to train the classifier (BioCOP2009-4 for gender and BioCOP2009-5 for race). Only subjects from the first of five random partitions (see Section 4.5) were used to train the classifier. The results are displayed in Table 4.6.

For the 5×6 image resolution, it can be observed that the cross-dataset gender prediction accuracy of 81.3% for the left NIR ocular images actually outperforms the 75.0 ± 1.0% prediction accuracy achieved by training and testing on images from the same dataset (see Figure 4.5 and Table 4.5). For race, the cross-dataset prediction accuracy for left NIR ocular images is 69.2%, which is ~ 9% lower than the same-dataset results of 77.9 ± 1.6% (see Figure 4.5 and Table 4.5).

The gender prediction accuracy for the right eye is noticeably less than the left eye at both image resolutions, while race prediction does not exhibit a similar trend. The relatively high prediction accuracy for gender was achieved while using images that *contained* cosmetic and contact lenses. This further reinforces the observation that gender cues are perhaps more prominent in the periocular region compared to the iris region.

4.5.5 CNN Optimization

The CNNs used in the previous experiments were kept as similar as possible across the different image resolutions. This was done so that the performance on the input images could be attributed to the image resolution as opposed to the CNN architecture. However, we were interested in

Table 4.6: Cross-dataset prediction accuracy (in %). A CNN model trained on the BioCOP2009 images and tested on the Cosmetic Contact dataset.

ImageSize	Laterality	Gender	Race
5×6	Left	81.3	69.2
	Right	72.1	72.5
10×12	Left	80.4	70.1
	Right	75.4	71.2

Table 4.7: CNN-based method. Attribute prediction accuracy (in %) for a CNN optimized for the 5×6 image input.

	Gender		Race		Eye Color	
Size	Left	Right	Left	Right	Left	Right
5×6 (Opt.)	77.1 ± 0.8	77.6 ± 1.4	84.0 ± 1.8	84.2 ± 0.9	77.6 ± 1.7	77.6 ± 1.2

determining if the performance using the 5×6 images could be further improved by judiciously modifying the associated CNN. Given the freedom to modify the hyperparameters, we were able to generate a CNN model that was only slightly better for gender and eye color prediction; however, race prediction did see a significant improvement in accuracy (see Table 4.7). The race prediction accuracy for the left ocular images increased by ~6% from $77.9 \pm 1.6\%$ to $84.0 \pm 1.5\%$ and for the right by ~8% from $76.5 \pm 1.2\%$ to $84.2 \pm 0.9\%$.

4.5.6 Special Case: Attribute Classification Based on a Single Pixel Image

The previous sections presented experiments where the image was down sampled to a resolution as small as 2×3, what if the near infrared ocular images were downsized to a single pixel? How much discriminatory attribute information would be available? In this section we will discuss the experiments performed in order to make this determination. Given texture in computer vision is typically measured by quantizing the relationship between neighboring pixels, the texture-based method utilizing the BSIF descriptor was not applied. The simple prototype based method and CNN-based method however were applied. For the simple prototype based method, the same protocol described in Section 4.5.1 was adhered to. For the CNN-based method, a convolutional layer for an input of a single pixel is not necessary and therefore only the fully connected layer was

Table 4.8: Special case: Attribute prediction from a single pixel.

Method	Gender		Race		Eye Color	
	Left	Right	Left	Right	Left	Right
Prototype-based	54.2 ± 0.8	54.9 ± 0.9	55.6 ± 0.8	56.1 ± 0.8	50.8 ± 1.2	51.4 ± 1.2
CNN-based	54.2 ± 0.8	55.0 ± 0.9	55.6 ± 0.8	56.2 ± 0.7	51.1 ± 1.4	51.8 ± 1.3

utilized (essentially a neural network classifier). In order to allow the network to learn nonlinear relationships an additional hidden layer was added, therefore the final architecture utilized is as follows: single pixel input layer, fully connected layer (2 nodes), fully connected layer (2 nodes), softmax layer, classification layer. The results utilizing a single input pixel for the simple prototype based method and the 'CNN-based' method are shown in Table 4.8.

4.6 Summary and Future Work

In this chapter, we conducted an experiment to determine the impact of image resolution on attribute prediction in the context of near-infrared ocular images. Two attribute prediction models were used for this purpose: a BSIF-based method and a CNN-based method. The CNN-based method resulted in the best prediction accuracy with $77.1 \pm 0.8\%$ for gender, $84.0 \pm 1.8\%$ for race, and $77.6 \pm 1.7\%$ for eye color, on images of size 5×6 for left ocular images, and $77.6 \pm 1.4\%$ for gender, $84.2 \pm 0.9\%$ for race, and $77.6 \pm 1.2\%$ for eye color on right ocular images.

The CNN-based method was also shown to generalize reasonably well when trained on one dataset and tested on another.

The observation that a 5×6 ocular image can be used for gender or race or eye color prediction is indeed very surprising. To be sure, the performance numbers for the three attributes considered in this chapter are below 85% at that resolution. Nevertheless, the drop in prediction accuracy for low resolution images is not as steep as we would expect (Figure 4.5 and Table 4.5). One explanation, in the context of CNNs, is that smaller networks (in terms of number of weights to be learned) require fewer training samples than larger networks. Perhaps this worked in favor of the low resolution images in this chapter (see [32]). Another possible explanation has to do with the limited number of classes being considered for each attribute (viz., 2). If the number of classes

were increased, prediction accuracies may plunge at a faster rate for low resolution images.

Notwithstanding these explanations, we must concede at this time that the precise cues being harnessed from low resolution images (i.e., 30 pixels worth of data) is not known. This will be the subject of a future study.

We would like to determine if more sophisticated methods can be developed to extract attributes from low resolution images. Further, attributes such as race and eye color should be considered to be multi-valued rather than binary. Even though it was shown that the CNN-based method generalized to an entirely different dataset captured with a *different* sensor in a *different* environment at a *different* location, the impact on other type of non-idealities, besides down-sampled images, should also be investigated.

CHAPTER 5

CROSS ATTRIBUTE PREDICTION

5.1 Introduction

Attribute prediction methods for NIR ocular images have been proposed using a texture-based method, a pixel intensity-based method and a Convolutional Neural Network based method. In this chapter, we explore the relationship between attributes as assessed by a CNN. This is accomplished by extracting a feature vector from the trained CNN model for a particular attribute (e.g., gender) and utilizing this feature vector to predict a different attribute. We refer to the feature vector as an 'attributeCode' (e.g, genderCode). The attributeCodes pertaining to a particular attribute are then used to train a support vector machine (SVM) in order to predict the other attribute. The feature vector itself corresponds to the activation of the last convolutional layer just before the fully connected layer. Experiments in this chapter will demonstrate the feasibility of cross attribute prediction. The remaining sections are organized as follows: related work is discussed in Section 5.2, the proposed feature extraction method in Section 5.3, the experimental results in Section 5.5 and a summary and future work in Section 5.6.

5.2 Related Work

The use a CNN for computer vision and pattern recognition problems has rapidly increased over the last several years. A CNN uses a large amount of data to train a model to learn features from the data. The features are learned (extracted) by the convolutional layers while the classification occurs uses the fully connected layers [64] (see Figure 5.1. The features extracted by a CNN for one problem, may also be useful for other problems [27]. Such an approach is referred to as transfer learning. Transfer learning is a method in which a model that is trained for one task (e.g., dog classification) is used for a different task (e.g., bird classification) by fine tuning some of the weights in the trained model.
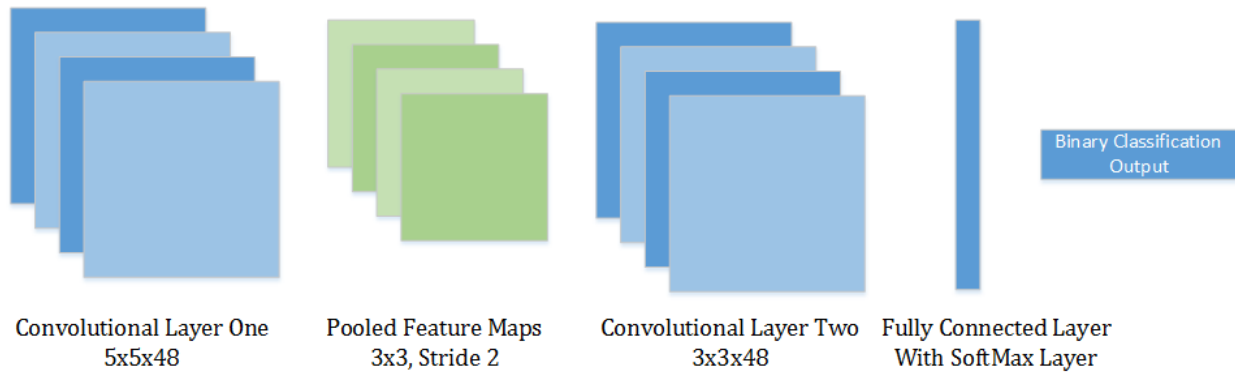
Figure 5.1: The Convolutional Neural Network architecture used in this chapter. The first layer consists of a convolutional layer with a 5×5 filter and 48 channels, then a relu layer, a max pooling layer (3×3 with a stride of 2), followed by a second convolutional layer with a 3×3 filter and 48 channels, followed by a relu layer, a fully connected layer, softmax layer and finally a binary classification layer.

Figure 5.2: Attribute code generation. An NIR ocular image is applied to the input of a Convolutional Neural Network trained to predict that attribute, the activation from the second convolutional layer is reshaped into a single one dimensional feature vector.

5.3   Feature Extraction

An NIR ocular image is input to a trained attribute prediction model and the activation of the second convolutional layer is used as a feature vector. This feature vector is refered to as a *gendercode* for the feature vector generated from a gender prediction CNN, and *racecode* for the feature vector generated from a race prediction CNN (see Figure 5.2). The experiments in this chapter utilize NIR ocular images of 42×50 resolution, though other image resolutions from Chapter 4 could be selected. The 42×50 image resolution was selected based on the tradeoff between attribute prediction accuracy and number of input pixels.

83

Table 5.1: The number of subjects in the BioCOP2009-4 dataset that were available for the experiments in Chapter 5. 125 subjects from each category were randomly selected to be used for the experiments.

Gender and Race Category	# of Subjects
Male Caucasian	332
Female Caucasian	504
Male Non-Caucasian	135
Female Non-Caucasian	125
Total Subjects	1096

5.4 Dataset

The BioCOP2009-4 dataset (see Section 2.5.1.4) was used for the experiments in this chapter. The subjects were categorized into 4 unique classes determined by their race and gender labels: a) Male Caucasian b) Female Caucasian c) Male non-Caucasian d) Female non-Caucasian. The number of subjects available in each category is displayed in Table 5.1. The subjects were further divided into three partitions: training, validation and test. A strict subject disjoint protocol was adhered too, meaning images of each subject were all placed in one partition only (no subject overlap between partitions). Training with balanced classes was desired; therefore the number of subjects in each class was fixed to be of the same size as the smallest class. The female non-Caucasian category was the smallest class of size 125 subjects. For each of the other three categories, 125 subjects were randomly selected to be included in the experiments. After the random selection of 125 subjects, 60% of the subjects from each of the four categories were randomly selected for training, 20% for validation and 20% for testing; i.e., 79, 23 and 23 subjects, respectively. Images from the subjects in the training partition were used for training purposes, images from the subjects in the validation model were used to optimize the trained model, while images from the subjects in the test partition were used to report the performance of the trained models.

5.5 Experiments

Two attribute prediction models were trained using a 2 layer CNN, the architecture is displayed in Figure 5.1. One attribute prediction model was trained to predict race, Caucasian or non-

Caucasian, while the other attribute prediction model was trained to predict gender, male or female. Each model was trained on images from subjects in the training partition. The validation set is used to optimize the trained model. The resulting attribute prediction accuracy on the images from the test subjects is displayed in Table 5.2. Each of the trained attribute prediction models is used to generate the following attribute codes:

GenderCode: An NIR ocular image is applied as input to the trained gender prediction model. The activation of the second convolutional layer is reshaped into a one dimensional vector, defined as the *genderCode*. For the CNN displayed in Figure 5.1, the activation is of size $17 \times 21 \times 48$, which is reshaped to a one dimensional feature vector of length $17,136$.

RaceCode: An NIR ocular image is applied as input to the trained race prediction model. The activation of the second convolutional layer is reshaped into a one dimensional vector, defined as the *raceCode*. For the CNN displayed in Figure 5.1, the activation is of size $17 \times 21 \times 48$, which is reshaped to a one dimensional feature vector of length $17,136$.

A genderCode for each image in the training partition was generated as described above, and a support vector machine (SVM) was trained using the genderCode and it's corresponding race label to generate a race prediction model. The trained SVM is used to predict race labels for each image in the test partition. The prediction accuracy is displayed in Table 5.3. The results indicate that race *can* be predicted, with less than 3% decrease in prediction accuracy (from 87.6% to 84.8%), from the features encoded in the genderCode. Race prediction as a function of race and gender is displayed in Table 5.5. The two methods agreed on correct Caucasian predictions 80.5% of the time and correct non-Caucasians predictions 81.1% of the time. The experimental results lead us to conclude that the exact same features are not being extracted by each method, though there may be a large number of similar features.

A raceCode for each image in the training partition was also generated, and an SVM was trained using the raceCode and it's corresponding gender label to generate a gender prediction model. The trained SVM model is used to predict gender labels for each image in the test partition. The prediction accuracy is displayed in Figure 5.3. The results indicate that utilizing the features

Table 5.2: The attribute prediction accuracy for each of the CNN models used to generate the attribute codes.

Attribute	Prediction Accuracy
Gender	79.4%
Race	87.6%

Table 5.3: The attribute prediction accuracy using an SVM to classify the attribute codes.

Predicted Attribute	Attribute Code	Prediction Accuracy
Gender	Race	79.2%
Race	Gender	84.8%

Table 5.4: Race prediction as a function of race and gender from both the CNN-based method and genderCode with SVM as a classifier.

Method	Images Correctly Predicted			
	Caucasian		Non-Caucasian	
	Male	Female	Male	Female
CNN	423 of 460	381 of 460	409 of 465	379 of 465
genderCode	408 of 460	368 of 460	403 of 465	386 of 465

encoded in the raceCode, gender can also be predicted with the same accuracy (from 79.4% to 79.2%, down only by 0.2%). Gender prediction as a function of race and gender is displayed in Table 5.4. The two methods agreed on correct male predictions 74.3% of the time and correct female predictions 69.4% of the time. The experimental results lead us to conclude that the same features are not being extracted by each method, though there may be a number of similar features.

Table 5.5: Gender prediction as a function of race and gender from the CNN-based method as well as from raceCode with SVM as a classifier.

Method	Images Correctly Predicted			
	Male		Female	
	Caucasian	Non-Caucasian	Caucasian	Non-Caucasian
CNN	388 of 460	368 of 460	374 of 460	335 of 465
raceCode	376 of 460	359 of 460	375 of 465	342 of 465

5.6 Summary and Future Work

In this chapter, we have introduced a novel problem, termed as cross attribute prediction, where feature vectors generated for predicting one attribute are directly used to predict a different attribute. It was shown that the genderCode, generated from a model trained to predict gender, is capable of predicting race with only a slight decrease in prediction accuracy (from 87.6% to 84.8%). It was also shown that the raceCode, generated from a model trained to predict race, is capable of predicting gender with almost the same prediction accuracy (down only 0.3%, from 79.4% to 79.1%). This indicates that features learned to be of discriminatory value for one attribute contain cues of other attributes as well.

Future work could be expanded to include additional attributes such as eye color or age. In addition, fusion models could be explored to combine multiple attribute codes in order to increase prediction accuracy beyond what was presented in this chapter.

CHAPTER 6

THESIS CONTRIBUTIONS AND FUTURE WORK

In Chapter 2, a state of the art method was presented utilizing BSIF as a texture descriptor and SVM as a classifier to predict *gender* and *race* from a NIR ocular image that does *not* require segmentation or normalization. The ocular region was also shown to provide greater gender prediction accuracy than the iris-only region. A covariate analysis was performed to reveal the impact of different variables on gender and race prediction. The impact of image blur on gender and race prediction also demonstrated that race prediction accuracy decreases at a much faster rate than that of gender as the level of image blur increases.

In Chapter 3, a state of the art method to predict *eye color* utilizing texture with BSIF as a texture descriptor and SVM as a classifier was presented. The method does not require segmentation or normalization of the NIR ocular image. The impact of gender and race on eye color was also provided.

In Chapter 4, it was shown that 5×6 images provided an attribute prediction accuracy similar to that of much larger image resolutions, such as 340×400, utilizing a simple CNN. This conveyed the possibility of extracting soft biometric information from low resolution ocular images. It also illustrated the limitations of a CNN when trained with limited amout of data.

In Chapter 5, cross-attribute prediction was explored, wherein the feature vector used to predict gender was harnessed by an SVM classifier to predict race (and vice-versa). The feature vectors themselves were gleaned from the activation layers of a gender prediction CNN and a race prediction CNN. The experiments in this chapter suggest that gender cues are available in the race feature vector, and vice-versa.

In addition to the aforementioned contributions, the research in this dissertation has brought to the forefront the importance of using a subject disjoint train and test protocol. If a subject disjoint train/test protocol is not followed, an optimistically biased classifier results. Prior to my first publication [8], the research in this field did not consistently follow a subject disjoint train/test

protocol.

Future work should explore the possibility of predicting multi-valued attributes beyond just binary-valued variables (e.g., age and multi-valued ethnicity). Analysis of the iris and ocular texture should be conducted to determine if certain texture patches correspond to specific attribute classes. Fusion of the information from the left and right eye images could be explored to determine if a fusion model is capable of increasing attribute prediction accuracy. Further, it is necessary to determine the types of anatomical features being extracted by BSIF and CNN from low-resolution ocular images for attribute prediction.

Another future work would be the use of attribute codes for the problem of iris recognition. Attribute-based person recognition from NIR ocular images has great potential to be an interesting area of research. The ability to predict attributes from low resolution images, as shown in this thesis, could in turn be applied to person recognition from low resolution images.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] A. H. Abdulnabi, G. Wang, J. Lu, and K. Jia. Multi-task cnn model for attribute prediction. *IEEE Transactions on Multimedia*, 17(11):1949–1959, 2015.

[2] A. Andreopoulos and J. K. Tsotsos. 50 years of object recognition: Directions forward. *Computer vision and image understanding*, 117(8):827–891, 2013.

[3] M. Ansari. *Atlas of ocular anatomy*. Springer, Switzerland, 2016.

[4] S. Baluja and H. A. Rowley. Boosting sex identification performance. *International Journal of computer vision*, 71(1):111–119, 2007.

[5] A. Bansal, R. Agarwal, and R. K. Sharma. SVM based gender classification using iris images. In *Proc. of International Conference on Computational Intelligence and Communication Networks (CICN)*, pages 425–429, Nov 2012.

[6] R. Barnard, V. Pauca, T. Torgersen, R. Plemmons, S. Prasad, J. van der Gracht, J. Nagy, J. Chung, G. Behrmann, S. Mathews, et al. High-resolution iris image reconstruction from low-resolution imagery. In *Advanced Signal Processing Algorithms, Architectures, and Implementations XVI*, volume 6313, page 63130D. International Society for Optics and Photonics, 2006.

[7] M. S. Billinger. Another look at ethnicity as a biological concept: Moving anthropology beyond the race concept. *Critique of Anthropology*, 27(1):5–35, 2007.

[8] D. Bobeldyk and A. Ross. Iris or periocular? Exploring sex prediction from near infrared ocular images. In *IEEE International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–7, 2016.

[9] D. Bobeldyk and A. Ross. Predicting eye color from near infrared iris images. In *IAPR International Conference on Biometrics (ICB)*, 2018.

[10] D. Bobeldyk and A. Ross. Predicting gender and race from near infrared iris and periocular images. *arXiv preprint arXiv:1805.01912*, 2018.

[11] C. Boyce, A. Ross, M. Monaco, L. Hornak, and X. Li. Multispectral iris analysis: A preliminary study. In *Computer Vision and Pattern Recognition Workshops*, pages 51–59, 2006.

[12] R. T. Chin and C. R. Dyer. Model-based recognition in robot vision. *ACM Computing Surveys (CSUR)*, 18(1):67–108, 1986.

[13] A. D. Clark, S. A. Kulp, I. H. Herron, and A. A. Ross. A theoretical model for describing iris dynamics. In *Handbook of Iris Recognition*, pages 129–150. Springer, 2013.

[14] M. Da Costa-Abreu, M. Fairhurst, and M. Erbilek. Exploring gender prediction from iris biometrics. In *IEEE International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–11, 2015.

[15] A. Dantcheva, P. Elia, and A. Ross. What else does your biometric data reveal? A survey on soft biometrics. *IEEE Transactions on Information Forensics And Security (TIFS)*, 11:441–467, 2016.

[16] A. Dantcheva, N. Erdogmus, and J.-L. Dugelay. On the reliability of eye color as a soft biometric trait. In *IEEE Workshop on Applications of Computer Vision (WACV)*, pages 227–231, 2011.

[17] J. Daugman. The importance of being random: statistical principles of iris recognition. *Pattern Recognition*, 36(2):279–291, 2003.

[18] J. Daugman. How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):21–30, 2004.

[19] H. Ding, D. Huang, Y. Wang, and L. Chen. Facial ethnicity classification based on boosted local texture and shape descriptions. In *10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, pages 1–6, 2013.

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

[21] J. S. Doyle, K. W. Bowyer, and P. J. Flynn. Variation in accuracy of textured contact lens detection based on sensor and lens pattern. In *Proc. of IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–7, 2013.

[22] M. Edwards, D. Cha, S. Krithika, M. Johnson, and E. J. Parra. Analysis of iris surface features in populations of diverse ancestry. *Open Science*, 3(1):150424, 2016.

[23] S. El-Naggar and A. Ross. Which dataset is this iris image from? In *IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6, Nov 2015.

[24] G. Fahmy. Super-resolution construction of iris images from a visual low resolution face video. In *9th International Symposium on Signal Processing and Its Applications*, pages 1–6. IEEE, 2007.

[25] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth. Describing objects by their attributes. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 1778–1785, 2009.

[26] B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski. Sexnet: A neural network identifies sex from human faces. In *NIPS*, volume 1, page 2, 1990.

[27] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.

[28] G. Guo. Human age estimation and sex classification. *Video Analytics for Business Intelligence*, 409:101–131, 2012.

[29] G. Guo, C. R. Dyer, Y. Fu, and T. S. Huang. Is gender recognition affected by age? In *IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 2032–2039, 2009.

[30] G. Guo, G. Mu, Y. Fu, C. Dyer, and T. Huang. A study on automatic age estimation using a large database. In *IEEE International Conference on Computer Vision*, pages 1986–1991, 2009.

[31] J. J. Howard and D. Etter. The effect of ethnicity, gender, eye color and wavelength on the biometric menagerie. In *IEEE International Conference on Technologies for Homeland Security (HST)*, pages 627–632, 2013.

[32] P. Hu and D. Ramanan. Finding tiny faces. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1522–1530. IEEE, 2017.

[33] J. Huang, L. Ma, T. Tan, and Y. Wang. Learning based resolution enhancement of iris images. In *British Machine Vision Conference*, pages 1–10, 2003.

[34] A. Hyvärinen, J. Hurri, and P. O. Hoyer. *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision*, volume 39. Springer, 2009.

[35] A. Jain, S. Dass, and K. Nandakumar. Can soft biometric traits assist user recognition? In *Defense and Security*, number 5404, pages 561–572. International Society for Optics and Photonics, April 2004.

[36] A. Jain, B. Klare, and A. Ross. Guidelines for best practices in biometrics research. In *IAPR International Conference on Biometrics (ICB)*, pages 541–545, 2015.

[37] A. K. Jain, A. A. Ross, and K. Nandakumar. *Introduction to biometrics*. Springer, New York, 2011.

[38] R. Jillela and A. Ross. Matching face against iris images using periocular information. In *IEEE International Conference on Image Processing (ICIP)*, pages 4997–5001, 2014.

[39] R. Jillela, A. Ross, and P. J. Flynn. Information fusion in low-resolution iris videos using principal components transform. In *IEEE Workshop on Applications of Computer Vision (WACV)*, pages 262–269, 2011.

[40] J. Kannala and E. Rahtu. BSIF: Binarized statistical image features. In *Proc. of International Conference on Pattern Recognition (ICPR)*, pages 1363–1366, 2012.

[41] H.-C. Kim, D. Kim, Z. Ghahramani, and S. Y. Bang. Appearance-based gender classification with gaussian processes. *Pattern Recognition Letters*, 27(6):618–626, 2006.

[42] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[43] A. Kuehlkamp, B. Becker, and K. Bowyer. Gender-from-iris or gender-from-mascara? In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1151–1159, 2017.

[44] A. Kuehlkamp and K. Bowyer. Predicting gender from iris texture may be harder than it seems. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 904–912. IEEE, 2019.

[45] N. Kumar, A. Berg, P. N. Belhumeur, and S. Nayar. Describable visual attributes for face verification and image search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(10):1962–1977, 2011.

[46] S. Lagree and K. W. Bowyer. Predicting ethnicity and gender from iris texture. In *IEEE International Conference on Technologies for Homeland Security (HST)*, pages 440–445, 2011.

[47] C. H. Lampert, H. Nickisch, and S. Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 951–958, 2009.

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

[49] L.-J. Li, H. Su, Y. Lim, and L. Fei-Fei. Objects as attributes for scene classification. In *European Conference on Computer Vision*, pages 57–69. Springer, 2010.

[50] X. Lu, H. Chen, and A. K. Jain. Multimodal facial gender and ethnicity identification. In *IAPR International Conference on Biometrics*, pages 554–561. Springer, 2006.

[51] J. R. Lyle, P. E. Miller, S. J. Pundlik, and D. L. Woodard. Soft biometric classification using periocular region features. In *Proc. of IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*, pages 1–7, 2010.

[52] J. Merkow, B. Jou, and M. Savvides. An exploration of gender identification using only the periocular region. In *Proc. of IEEE Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–5, 2010.

[53] V. Mirjalili, S. Raschka, and A. Ross. Gender privacy: An ensemble of semi adversarial networks for confounding arbitrary gender classifiers. In *9th IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2018.

[54] S. Nagabhyru. *Gender Estimation from Fingerprints Using DWT and Entropy*. PhD thesis, West Virginia University, Morgantown, 2016.

[55] K. Nguyen, C. Fookes, R. Jillela, S. Sridharan, and A. Ross. Long range iris recognition: A survey. *Pattern Recognition*, 72:123–143, 2017.

[56] K. Nguyen, C. Fookes, S. Sridharan, and S. Denman. Feature-domain super-resolution for iris recognition. *Computer Vision and Image Understanding*, 117(10):1526–1535, 2013.

[57] K. Nguyen, C. Fookes, S. Sridharan, M. Tistarelli, and M. Nixon. Super-resolution for biometrics: A comprehensive survey. *Pattern Recognition*, 78:23–42, 2018.

[58] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.

[59] V. Ojansivu and J. Heikkilä. Blur insensitive texture classification using local phase quantization. In *International conference on image and signal processing*, pages 236–243. Springer, 2008.

[60] X. Qiu, Z. Sun, and T. Tan. Global texture analysis of iris images for ethnic classification. In *IAPR International Conference on Biometrics*, pages 411–418. Springer, 2006.

[61] X. Qiu, Z. Sun, and T. Tan. Learning appearance primitives of iris images for ethnic classification. In *Proc. of IEEE International Conference on Image Processing (ICIP)*, volume 2, pages II–405, 2007.

[62] R. Raghavendra and C. Busch. Robust scheme for iris presentation attack detection using multiscale binarized statistical image features. *IEEE Transactions on Information Forensics and Security*, 10(4):703–715, 2015.

[63] E. Ramón-Balmaseda, J. Lorenzo-Navarro, and M. Castrillón-Santana. Gender classification in large databases. In *Iberoamerican Congress on Pattern Recognition*, pages 74–81. Springer, 2012.

[64] S. Raschka and V. Mirjalili. *Python Machine Learning, 2nd Ed.* Packt Publishing, Birmingham, UK, 2017.

[65] A. Rattani, N. Reddy, and R. Derakhshani. Gender prediction from mobile ocular images: A feasibility study. In *IEEE International Symposium on Technologies for Homeland Security (HST)*, pages 1–6, 2017.

[66] D. K. Roberts, A. Lukic, Y. Yang, J. T. Wilensky, and M. N. Wernick. Multispectral diagnostic imaging of the iris in pigment dispersion syndrome. *Journal of glaucoma*, 21(6):351–357, 2012.

[67] A. Ross and C. Chen. Can gender be predicted from near-infrared face images? *Image analysis and recognition*, pages 120–129, 2011.

[68] J. A. Sanchis-Gimeno, D. Sanchez-Zuriaga, and F. Martinez-Soriano. White-to-white corneal diameter, pupil diameter, central corneal thickness and thinnest corneal thickness values of emmetropic subjects. *Surgical and radiologic anatomy*, 34(2):167–170, 2012.

[69] H. J. Santos-Villalobos, D. R. Barstow, M. Karakaya, C. B. Boehnen, and E. Chaum. Ornl biometric eye model for iris recognition. In *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 176–182, 2012.

[70] W. J. Scheirer, N. Kumar, V. N. Iyer, P. N. Belhumeur, and T. E. Boult. How reliable are your visual attributes? In *Biometric and Surveillance Technology for Human and Activity Identification X*, volume 8712, page 87120Q. International Society for Optics and Photonics, 2013.

[71] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[72] M. Singh, S. Nagpal, M. Vatsa, R. Singh, A. Noore, and A. Majumdar. Gender and ethnicity classification of iris images using deep class encoder. In *Proc. of IEEE International Joint Conference on Biometrics (IJCB)*, 2017.

[73] M. Singh, S. Nagpal, M. Vatsa, R. Singh, A. Noore, and A. Majumdar. Gender and ethnicity classification of iris images using deep class-encoder. In *IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–8, October 2017.

[74] R. A. Sturm and M. Larsson. Genetics of human iris colour and patterns. *Pigment cell & melanoma research*, 22(5):544–562, 2009.

[75] N. Sun, W. Zheng, C. Sun, C. Zou, and L. Zhao. Gender classification based on boosting local binary pattern. In *International Symposium on Neural Networks*, pages 194–201. Springer, 2006.

[76] T. Suzuki, S. M. Richards, S. Liu, R. V. Jensen, and D. A. Sullivan. Influence of sex on gene expression in human corneal epithelial cells. *Investigative Ophthalmology & Visual Science*, 47(13):1584–1584, 2006.

[77] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *AAAI*, volume 4, page 12, 2017.

[78] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[79] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[80] J. Tapia and C. Aravena. Gender classification from NIR iris images using deep learning. In *Deep Learning for Biometrics*, pages 219–239. Springer, 2017.

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

[82] J. Tapia and I. Viedma. Gender classification from multispectral periocular images. In *IEEE International Joint Conference on Biometrics (IJCB)*, pages 805–812, 2017.

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

[84] J. E. Tapia, C. A. Perez, and K. W. Bowyer. Gender classification from iris images using fusion of uniform local binary patterns. In *Proc. of ECCV Workshops*, pages 751–763. Springer, 2014.

[85] J. E. Tapia, C. A. Perez, and K. W. Bowyer. Gender classification from the same iris code used for recognition. *IEEE Transactions on Information Forensics and Security*, 11(8):1760–1770, 2016.

[86] V. Thomas, N. V. Chawla, K. W. Bowyer, and P. J. Flynn. Learning to predict gender from iris images. In *Proc. of IEEE Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–5, 2007.

[87] M. Toews and T. Arbel. Detection, localization, and sex classification of faces from arbitrary viewpoints and under occlusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9):1567–1581, 2009.

[88] B. N. Torgrimson and C. T. Minson. Sex and gender: what is the difference? *Applied Physiology*, 99(3):785–787, 2005.

[89] H. Wagner, B. A. Fink, and K. Zadnik. Sex-and gender-based differences in healthy and diseased eyes. *Optometry-Journal of the American Optometric Association*, 79(11):636–652, 2008.

[90] J.-G. Wang, J. Li, W.-Y. Yau, and E. Sung. Boosting dense sift descriptors and shape contexts of face images for gender recognition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 96–102, 2010.

[91] X. Wang and Q. Ji. Object recognition with hidden attributes. In *IJCAI*, pages 3498–3504, 2016.

[92] C. L. Wilkerson, N. A. Syed, M. R. Fisher, N. L. Robinson, D. M. Albert, et al. Melanocytes and iris color: light microscopic findings. *Archives of Ophthalmology*, 114(4):437–442, 1996.

[93] J.-H. Yoo, D. Hwang, and M. S. Nixon. Gender classification in human gait using support vector machine. In *ACIVS*, volume 5, pages 138–145. Springer, 2005.

[94] G. Zhang and Y. Wang. Multimodal 2D and 3D facial ethnicity classification. In *IEEE Fifth International Conference on Image and Graphics (ICIG)*, pages 928–932, 2009.