

Predicting Eye Color from Near Infrared Iris Images

Denton Bobeldyk^{1,2} Arun Ross¹

denny@bobeldyk.org rossarun@cse.msu.edu

¹ Michigan State University, East Lansing, USA

² Davenport University, Grand Rapids, USA

Abstract

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 work, 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.

1. Introduction

Iris recognition systems utilize the iris patterns evident in the eye for automated recognition of individuals [11]. A typical iris recognition system acquires the ocular image of an individual; segments the annular iris region from the input ocular image; unwraps and normalizes the annular iris region into a rectangular entity using a rubber sheet model; applies a set of Gabor filters to extract textural details from the normalized iris; binarizes the ensuing phase responses into a binary iriscodes; and determines the degree of dissimilarity between two iriscodes based on their Hamming Distance [6].

The iris is typically imaged in the near-infrared (NIR) spectrum (as opposed to the visible spectrum which pro-

duces 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 [3]; 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 [2]. 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 1.

It may seem implausible if not impossible to predict the ‘eye color’¹ of an individual based on 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.

1.1. 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 eye color. Darker color irides contain more melanin than lighter color irides [21]. 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 [21]. 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

¹Perceived eye color is perhaps a more accurate term, as the color of an individual’s eye can 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.

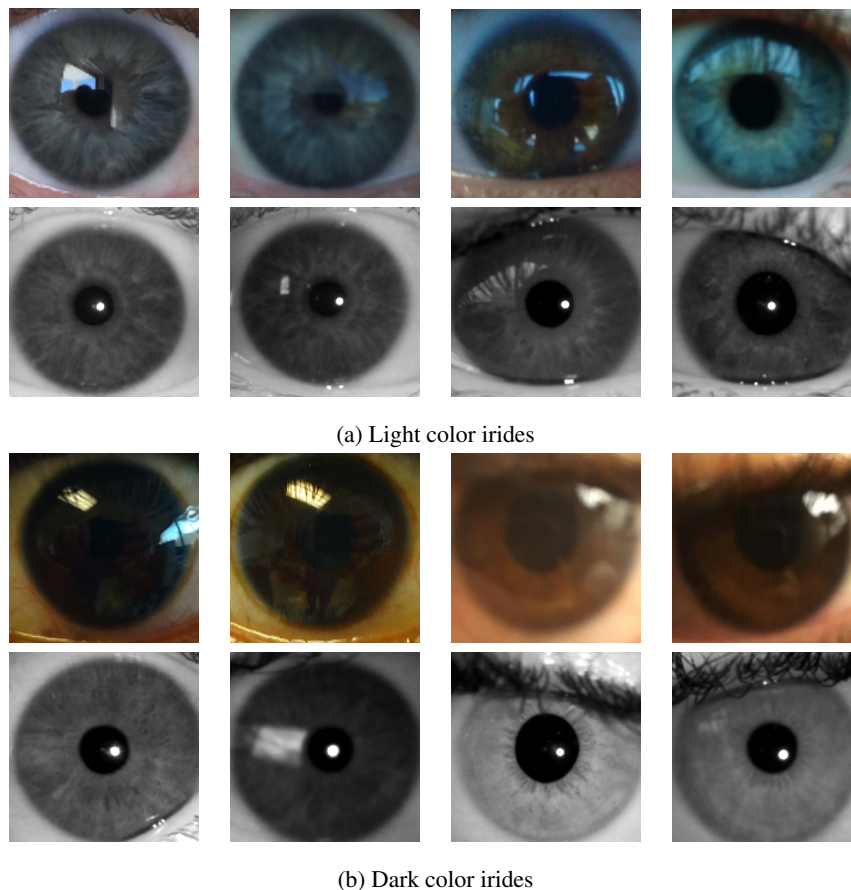


Figure 1: Examples of (a) light color irides, and (b) dark color irides. In each case, the top row shows images in the RGB color space and the bottom row shows the corresponding images in the NIR spectrum. The NIR images were taken with the Iritech IrisShield USB sensor while the RGB images were taken 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.

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’ [19]. This effect will make 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 [13] as well as soft biometric prediction from NIR iris images [1]. The BSIF descriptor has also shown success in other iris biometric problems such as presentation attack detection [7, 16].

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 [4]; (c) Eye color can also be used in cross-spectral matching scenarios, when comparing NIR iris images against RGB images [12]; (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 [17] 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’).

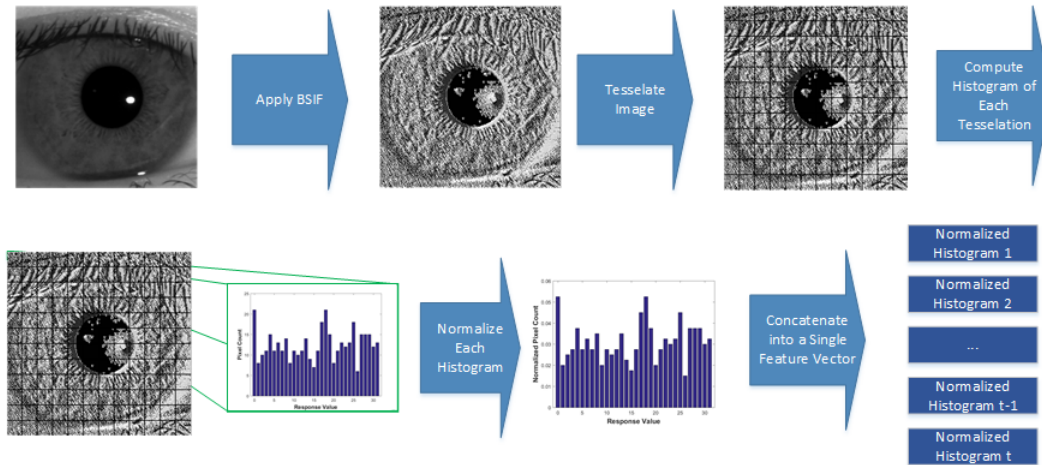


Figure 2: Generating the feature vector for eye color classification based on BSIF.

In this paper, we will refer to eye images labeled² with the color ‘brown’ as category A, and eye images labeled as ‘blue’, ‘green’, ‘hazel’, or ‘gray’ as category B. The rest of the paper is organized as follows: Section 2 discusses related work; Section 3 presents the two feature extraction methods used to predict eye color; Section 4 presents the dataset used; Section 5 presents the experiments and their results; Section 6 summarizes the findings of this work as well as discusses future work.

2. 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. [5] 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 [9] 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

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

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. Feature Extraction for Eye Color Prediction

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 [19, 21]. 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 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.

3.1. 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 [1, 20]. 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 [1] and the texture classification domain [13]. Due to this reason, the BSIF descriptor was used in this work.

The BSIF descriptor was introduced by Kanala and Rahtu [13]. BSIF projects the input image into a sub-

Table 1: Summary of the BioCOP 2009 dataset used in this work

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

Table 2: Eye Color, ethnicity and gender statistics of the BioCOP 2009 dataset.

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

space by convolving it with pre-generated filters. The pre-generated filters are created from 13 natural images supplied by the authors of [10]. 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 $[0, 2^n - 1]$.³

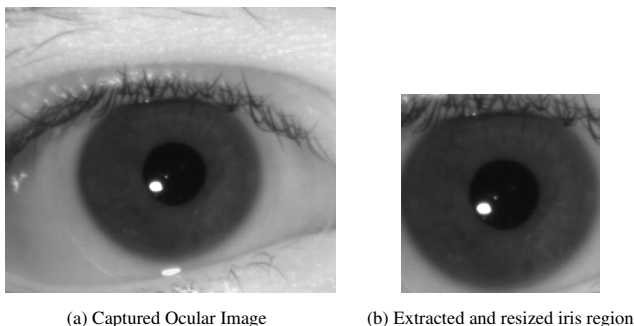


Figure 3: The iris region is extracted from the ocular image captured by the NIR sensor. Image taken from [8].

³While [13] states that the BSIF response is in the interval $[0, 2^n - 1]$, the matlab code supplied by the authors utilizes a range of $[1, 2^n]$

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 4 for details and Figure 3 for an example). The proposed texture-based method applies the BSIF operator to each NIR iris image. The filtered image is then tessellated 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 $n = 10$ and $k = 7$. These parameter values were selected empirically based on [1]. 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.

3.2. Intensity-based Method

In order to generate a feature vector based on pixel intensity, each iris image was once again tessellated 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 1 provides visual evidence that this is *not* the case, it is worth

Table 3: Number of images for each color category and label of the BioCOP09 dataset

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

Table 4: 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

confirming this in a rigorous manner.

4. BioCOP2009 Dataset

The BioCOP2009 dataset contains 43,810 NIR ocular images captured with 3 different iris sensors: 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 captured images of size 480×480 . Using a commercially available SDK, 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 (see Table 1).

The BioCOP2009 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 3. 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.

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 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 for training in category B were placed in the test partition; therefore each test set had a larger number of category B subjects than category A subjects. Table 4 summarizes the subject statistics of the experimental protocol adopted in this work.

In the iris recognition literature, differences in matching performance between left and right eyes have been observed [14, 15, 18]. This led us to conduct experiments separately on left and right eyes to determine if eye laterality had any impact on prediction accuracy.

5.1. Texture-based Method

The feature vectors that were generated using the texture-based method (see Subsection 3.1) were randomly partitioned by subject into 60% training and 40% testing as described above. The training feature vectors were used to generate an SVM classifier (with 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 5. The resulting confusion matrices for the left and right eye images are shown in Table 6.

5.2. Intensity-based method

The feature vectors that were generated from the intensity-based method (see Subsection 3.2) were randomly

Table 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 6: 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 7: 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

Table 8: 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
	Non Caucasian	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
	Non Caucasian	87.7 ± 1.3	87.4 ± 1.0

partitioned by subject into 60% training and 40% testing as described earlier. The training feature vectors were used to generate an SVM classifier (with 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 7. The overall classification accuracy is shown in Table 5.

5.3. Discussion

The prediction accuracy of the texture-based method outperforms that of the intensity-based method by 10% (see Table 5). This suggests that the intensity of NIR iris images cannot be solely used to predict eye color. Table 8 summarizes the results as a function of gender and ethnicity. 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). Iris images from Non Caucasian subjects were found to have a much higher prediction accuracy than the iris images from Caucasians; there was about a 6% difference using the texture-based method and about an 8% difference using the intensity-based method. We speculate this may be related to the higher number of Non Caucasian subjects in category A.

6. Summary and Future Work

The focus of this work 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 $\sim 90\%$ 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.

The proposed texture-based method could be expanded to not only predict between category A and B eye colors, but also to predict individual eye colors in category B {blue, green, hazel, gray}. It may be possible to discover anatomical differences between various categories of lighter color irides which could then be exploited to provide accurate prediction. The use of deep learning techniques or other texture descriptors (such as LBP, LPQ, etc.), in conjunction with BSIF, may be necessary to facilitate this.

7. Acknowledgements

This work was partially supported by the NSF Center for Identification Technology Research at West Virginia University.

References

- [1] 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.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 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. IEEE, 2011.
- [6] J. Daugman. How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):21–30, 2004.
- [7] J. S. Doyle and K. W. Bowyer. Robust detection of textured contact lenses in iris recognition using BSIF. *IEEE Access*, 3:1672–1683, 2015.
- [8] 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.
- [9] 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.
- [10] A. Hyvärinen, J. Hurri, and P. O. Hoyer. *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision*, volume 39. Springer Science & Business Media, 2009.
- [11] A. K. Jain, A. A. Ross, and K. Nandakumar. *Introduction to biometrics*. Springer, New York, 2011.
- [12] 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.
- [13] J. Kannala and E. Rahtu. BSIF: Binarized statistical image features. In *Proc. of International Conference on Pattern Recognition (ICPR)*, pages 1363–1366, 2012.
- [14] P. J. Phillips, K. W. Bowyer, P. J. Flynn, X. Liu, and W. T. Scruggs. The iris challenge evaluation 2005. In *Proc. of IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, pages 1–8, 2008.
- [15] P. J. Phillips, W. T. Scruggs, A. J. OToole, P. J. Flynn, K. W. Bowyer, C. L. Schott, and M. Sharpe. Frvt 2006 and ice 2006 large-scale results. *National Institute of Standards and Technology, NISTIR*, 7408(1), 2007.
- [16] 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.
- [17] 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.
- [18] A. SgROI, K. W. Bowyer, and P. Flynn. Effects of dominance and laterality on iris recognition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 52–58, 2012.
- [19] R. A. Sturm and M. Larsson. Genetics of human iris colour and patterns. *Pigment cell & melanoma research*, 22(5):544–562, 2009.
- [20] 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.
- [21] 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.