

Gender-From-Iris or Gender-From-Mascara?

Andrey Kuehlkamp Benedict Becker Kevin Bowyer
 Department of Computer Science & Engineering, University of Notre Dame
 akuehlka@nd.edu, kwb@nd.edu

Abstract

Predicting a person's gender based on the iris texture has been explored by several researchers. This paper considers several dimensions of experimental work on this problem, including person-disjoint train and test, and the effect of cosmetics on eyelash occlusion and imperfect segmentation. We also consider the use of multi-layer perceptron and convolutional neural networks as classifiers, comparing the use of data-driven and hand-crafted features. Our results suggest that the gender-from-iris problem is more difficult than has so far been appreciated. Estimating accuracy using a mean of N person-disjoint train and test partitions, and considering the effect of makeup - a combination of experimental conditions not present in any previous work - we find a much weaker ability to predict gender-from-iris texture than has been suggested in previous work.

1. Introduction

Classifying gender based on iris texture has been explored by several researchers, with a range of reported accuracies. Different features, classifiers and methods to evaluate accuracy have been used. Although the results indicate that the iris texture contains information related to gender, no work to date has described the texture appearance that characterizes each gender.

Neural Networks (NNs) are known as powerful classifiers, and for being able to autonomously learn features from the training data. Due to these properties, and to the current popularity of NN solutions in computer vision and biometrics, we explore the use of NNs for gender-from-iris. Apart from the classifier, several ways of extracting image features can be used. The simplest is to use pixel intensities, but more sophisticated techniques may result in more powerful features. We categorize feature extraction techniques into *data-driven*, which are learned automatically by the NN classifiers, and *hand-crafted*, that applies some specifically defined transformation over the raw data.

Most gender-from-iris work to date has overlooked one

or more questions that may be important: *What is the accuracy breakdown by gender? Is gender-from-iris based on true iris texture differences, or based on incidental factors such as presence/absence of eye makeup? How important is subject-disjoint training and testing in getting true performance estimates? Do Convolutional Neural Networks (CNNs) offer any performance improvement over hand-crafted features and classifiers for gender-from-iris?*

This paper describes results of experiments that explore these questions. We compare the use of Multi-Layer Perceptrons (MLPs) and CNNs for gender-from-iris. We use different approaches to extract information from the iris texture; we analyze the accuracy achieved for each gender; we look into the bias that may be created by the use of cosmetics; and we look at the bias that results from not using a subject-disjoint training and testing.

2. Related Works

The extraction of ancillary information from biometric traits is known as *soft biometrics*. As defined by Dantcheva *et al.* [2], "[s]oft biometric traits are physical, behavioral, or material accessories, which are associated with an individual, and which can be useful for recognizing an individual."

Gender is one soft biometric attribute, and gender recognition has been explored using biometric traits such as faces, fingerprints, gait and irises. The earliest work on gender-from-iris [15] used a classifier based on decision trees, and reported an accuracy of about 75%. They extracted hand-crafted geometric and texture features from log-Gabor-filtered images in a dataset of over 57,000 images. The training and testing sets were not person-disjoint, which typically results in a higher estimated accuracy than can be expected for new persons.

Later, [8] used a Support Vector Machine (SVM) classifier with features extracted using spot and line detectors and Law's texture measures. They used a dataset of 600 images and a cross-validation protocol with 2, 5 and 10 folds, with person-disjoint partitions. They considered both race-from-iris and gender-from-iris, and their classification accuracy on gender-from-iris ranged from 47% to 62%. A similar approach was used by [1], which used 2D Discrete Wavelet

Authors	Classifier	Accuracy(%)	Features	Dataset Size	Person-Disjoint	Cross-Validation	Cosmetics	Breakdown by Gender
Thomas <i>et al.</i> (2007) [15]	Decision Tree	75	Gabor filtering + Hand-Crafted	57,137	No	10f	No	No
Lagree <i>et al.</i> (2011) [8]	SVM	47.67–62.17	Gabor filtering + Hand-Crafted	600	Yes	2, 5 and 10f	No	No
Bansal <i>et al.</i> (2012) [1]	SVM	83.06	Hand-crafted + DWT	300	No	10f	No	No
Tapia <i>et al.</i> (2014) [13]	SVM	96.67	LBP	3,000	No	80/20	No	Yes
Fairhurst <i>et al.</i> (2015) [6]	Various (individual and combined)	49.61–89.74	Geometric+Texture Hand-Crafted	1,600	Yes	72/28	No	No
Tapia <i>et al.</i> (2016) [14]	SVM	91; 85.33	IrisCode	3,000; 3,000 ¹	No; Yes	80/20	No	Yes
This paper	MLP/CNN	66 ± 2.7	Intensity, Gabor filtering, LBP	3,000	Yes	80/20, 10x	Yes	Yes

Table 1: Overview of gender prediction from iris images.

Transform (DWT) in combination with hand-crafted statistical features to extract texture information from the images. Using an SVM to classify the irises, they reported accuracy up to 83% on a small dataset of 300 images.

In the work of [13], using an SVM to classify Local Binary Pattern (LBP) features extracted from 3,000 iris images yielded an accuracy of 96.67%. This was for an 80/20 *train/test* split, on non-person-disjoint partitions. The same authors used a similar technique to perform gender classification based on the IrisCode used for identification in [14]. In this work, they performed evaluation on two different datasets: one was person-disjoint, while the other was not, and the reported accuracy changed considerably. The person-disjoint dataset, called the Gender-from-Iris (GFI) dataset, is available to the research community.

In another study, [6] used an SVM in a combined consensus with other classifiers to achieve 81% accuracy on a person-disjoint dataset. They used a combination of geometric and texture features, selected via statistical methods, and a 72/28 training/testing split to prevent overfitting.

An overview of the techniques and results used so far is presented in Table 1. None of these works has looked systematically at the effect of cosmetics on accuracy of predicting gender-from-iris. Most of the works do not use a subject-disjoint training and testing, especially those reporting the highest accuracy. And these works report accuracy from a single random split into train-test data, rather than a mean of N random splits. Apart from [6], no other research employed neural networks for this task.

¹The first set of 3,000 was not person-disjoint, so the authors used another 3,000 images person-disjoint set.

3. Methods

We use the "Gender from Iris" (GFI) dataset ² used in [14], which to our knowledge is the only publicly available dataset for this problem. It consists of 1,500 left-eye and 1,500 right-eye images, for 3,000 total, representing 750 male and 750 female subjects. The 480 × 640, near-infrared images were obtained with an LG 4000 iris sensor.

Previous work generally reported accuracy based on a single random split of the data into train and test. The problem with this is that a single partitioning of the data into train and test can easily give an "optimistic" estimate of true accuracy. For this reason, in our experiments, a basic trial is a random 80/20 split into train and test data, and reported accuracy is averaged over ten trials. Each trial is person-disjoint training and testing. With this approach, we expect to obtain a more true estimate of accuracy.

The iris images were processed using IrisBee [9] to segment and normalize the iris region. Normalized iris images were stored in different resolutions: 40 × 240, 20 × 240, 10 × 240, 5 × 120, 3 × 60 and 2 × 30 pixels. As a result of the segmentation, a mask is generated for each image, marking where the iris texture is occluded, usually by eyelids or eyelashes. In the experiments that used raw pixel intensities as the features, the normalized iris images were used as feature inputs of the classifier. The sizes of the feature vectors were then 4800, 2400, 600, 180 and 60, respectively.

After performing training on a portion of the images, we use the test set to perform the evaluation, based on a simple criterion: given an unlabeled normalized iris, can we correctly predict the subject's gender? Two main feature extraction techniques were explored: data-driven features us-

²<https://sites.google.com/a/nd.edu/public-cvrl/data-sets>

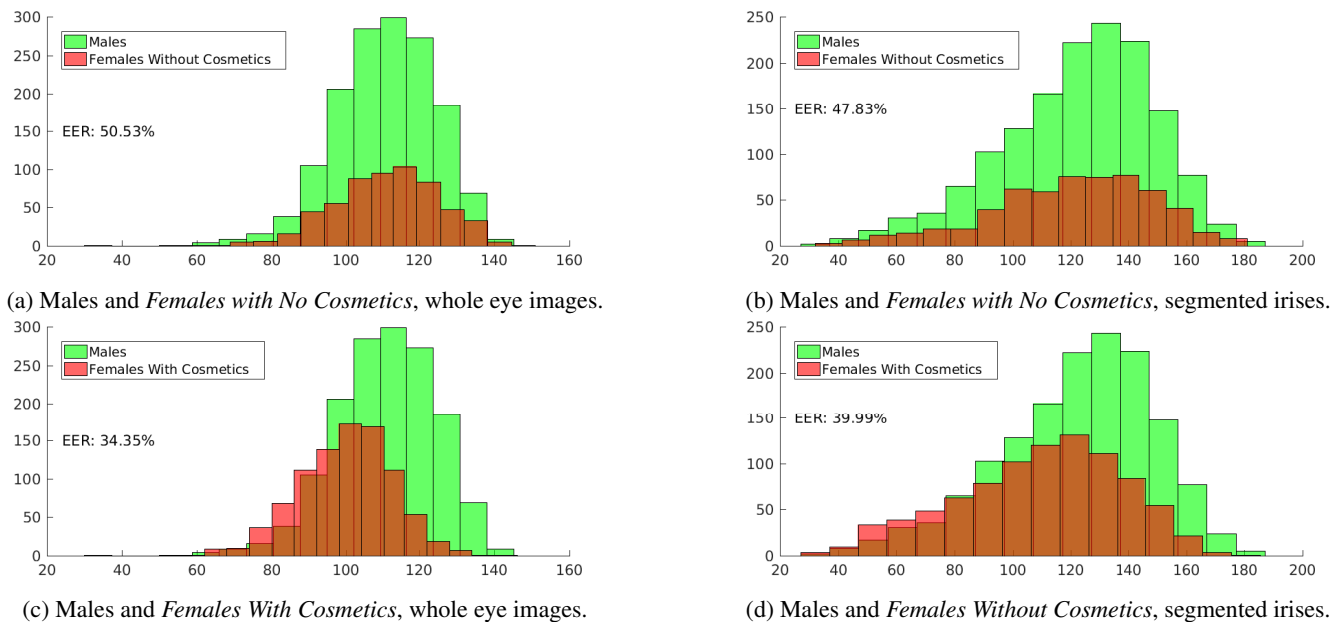


Figure 1: Threshold on average image intensity can achieve 60% correct classification of males and FWC.

ing raw pixel intensity, and hand-crafted features using Gabor filtering and LBP. A more detailed description of these feature extraction approaches is given in section 7.

Classification experiments were performed using MLP neural networks and CNNs. The details about the topology of the networks are described in section 8.

4. Person-Disjoint Train and Test

We performed the same experiment on the person-disjoint GFI dataset, and on a previous version of that dataset that is not person-disjoint. For the GFI dataset, there is one image per iris, and so the training and testing is necessarily person-disjoint. For the second dataset, there are a varying number of images per iris, of a smaller number of different irises, and so the training and testing is not person-disjoint. For both sets of results, accuracy is averaged over 10 trials, with each trial using a random 80/20 split for train/test data.

The estimated accuracy using the subject-disjoint training and testing enforced by the GFI dataset is $61\% \pm 2.9$. The estimated accuracy with the non-person-disjoint training and testing allowed by the other dataset with multiple images per iris is $77\% \pm 2.6$. This is an average over ten trials; Figure 2 shows that a single non-person-disjoint trial could easily result in an estimated accuracy of 100%. The higher estimated accuracy for the non-person-disjoint train/test apparently results from the classifier learning subject-specific features, rather than generic gender-related texture features.

This experiment makes the point that it is impossible

to meaningfully compare non-person-disjoint results with person-disjoint results. Higher (but optimistic) accuracies are reported for works using a non-subject-disjoint methodology and lower (but more realistic) accuracies reported using a subject-disjoint methodology. Also, in general, accuracies are reported for a single split of the data. A more useful accuracy estimate is computed over N trials using random person-disjoint splits of the data.

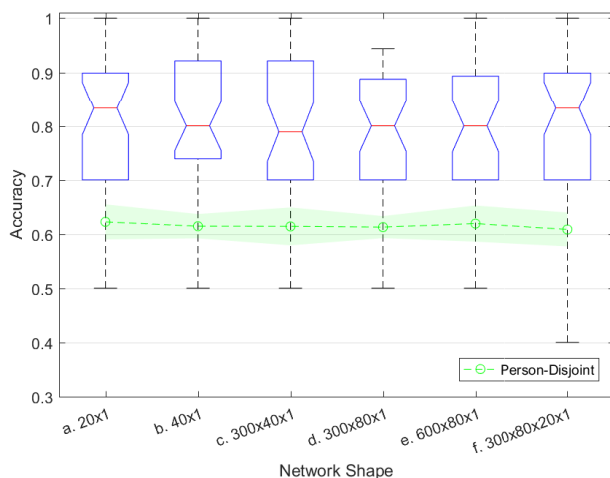


Figure 2: Box plots show accuracy distributions on a non-person-disjoint dataset, using an MLP classifier operating on raw pixel intensity. The green dotted line and shaded area represent the average accuracy and deviation on a person-disjoint dataset.

5. Male/Female or Mascara/No Mascara?

Mascara causes the eyelashes to appear thicker and darker in the iris image. Figure 3 shows a female eye with and without mascara. The use of eye makeup has been shown to affect iris recognition accuracy [5]. The basic mechanism is that if eyelash segmentation is not perfect, the segmented iris region may include some eyelash occlusion. To the degree that eyelash occlusion is present in the iris region, the use of mascara will generally increase the magnitude of the artifact in the texture computation. The same effect can also happen with other types of makeup like eyeliner, although this one is applied to the eyelid instead of the eyelashes.

To investigate how mascara might affect gender-from-iris results, we reviewed the GFI dataset and annotated which images show evidence of mascara or eyeliner. Just over 60% for the female iris images show visible evidence of cosmetics, compared to 0% of the male iris images. The annotation allowed us to perform experiments using three categories of images: Male, Female With Cosmetics (FWC) and Female with No Cosmetics (FNC).

One simple observation is that average image intensity for FWC is darker than for FNC or for Males (Fig. 1). This is true whether one considers the image as a whole, or only the segmented iris region. For Males and FNC, the distributions of average image intensity are almost identical; see Fig. 1a and 1b. For Males and FWC, there is a noticeable separation between the distributions; see Fig. 1c and 1d. Based on this separation, we could apply a simple threshold and achieve better than 60% accuracy distinguishing Males from FWC (EER of about 37%). However, a similar threshold for Males and FNC results in only about 50% accuracy. This experiment shows how the presence of mascara can potentially make the gender-from-iris problem appear to be easier to solve than it is in reality.

We also trained MLP networks to classify gender-from-iris. We considered both using the whole iris image, and using only the normalized iris region. We also considered training with and without images containing mascara. The results are summarized in Figure 4. When training with the full dataset (Males, FNC and FWC), the accuracy achieved with the whole image is greater than the accuracy achieved with the iris region alone. Also, the accuracy achieved is highest for the FWC subgroup, and lowest for the FNC subgroup. The trained MLP is apparently able to use the presence of mascara to correctly classify a higher fraction of the females in the FWC subgroup, at the expense of lower classification for the FNC subgroup.

Next we trained two additional networks, one using Males + FNC, and another using Males + FWC. The Male images were randomly sampled to equal the number of female images, to avoid biasing the training toward a majority class. Comparing the results for normalized iris

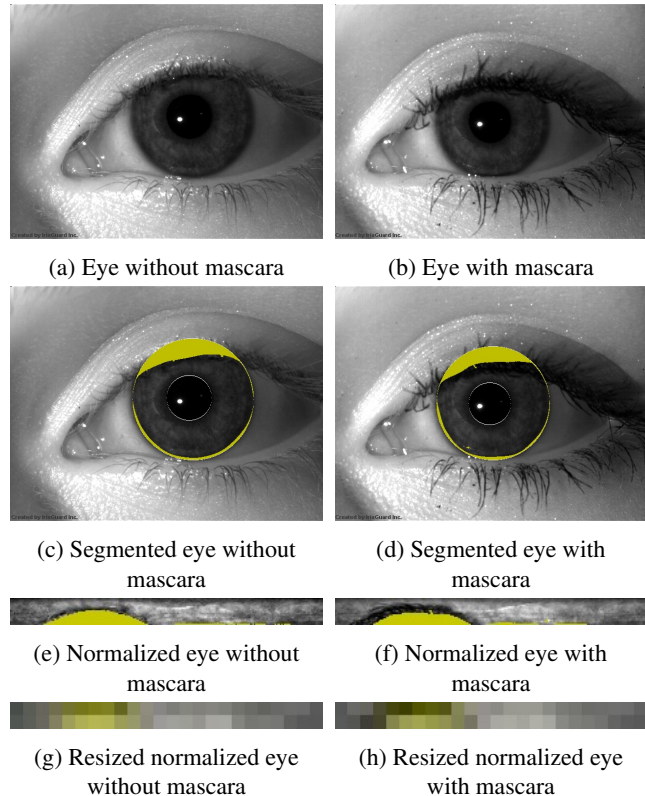


Figure 3: Segmentation and normalization process in two images of the same eye, with and without mascara.

trained on all subjects (Fig. 4a right) with those trained on Males+FNC (Fig. 4b right), while FNC performance improved, we can perceive a small decrease in the overall accuracy. At the same time, training on Males+FWC (Fig. 4c right) causes the overall performance to increase to 64%.

This effect is amplified when working with whole eye images. In Fig. 4b (left side) the FNC accuracy improvement is almost equal to the male accuracy drop, and it results in an overall accuracy contraction with regard to Fig. 4a (left). On the other hand, in Fig. 4c (left) the overall accuracy rises to 88%.

The experiment makes it clear that mascara is an important confounding factor for gender-from-iris. If mascara is present in the dataset, then it is hard to know the degree to which the classifier learns gender from iris texture versus gender from mascara. Future research on gender-from-iris should use datasets that include annotations for the presence of mascara, and new mascara-free datasets are needed.

6. Occlusion masks

Eyelids and eyelashes frequently occlude portions of the iris. Ideally, the segmentation step would result in these occlusions becoming part of the "mask" for the image. Re-

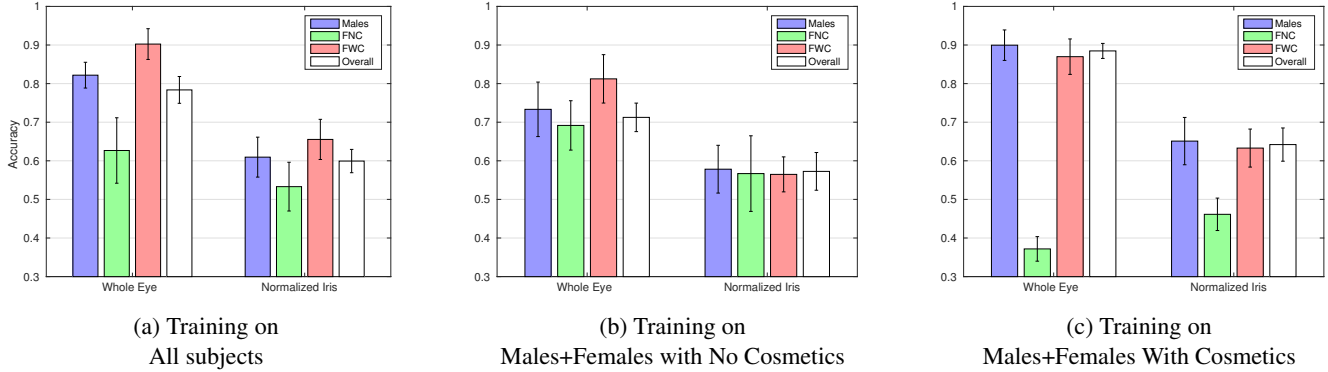


Figure 4: Gender classification accuracy for different training groups, using pixel intensity.

sults in the previous section indicate that eyelash occlusion is generally not perfectly segmented. It appears that mascara causes the "noise" resulting from un-masked eyelash occlusion to become a feature that can be correlated with gender. If this is the case, mascara may also cause more eyelash occlusion to be identified and segmented (Fig. 3). In this case, the size and shape of the masked region would be a feature correlated with gender.

In order to determine if the occlusion mask contains gender-related information, we performed an experiment where the only information given to the MLP classifier is the (binary) occlusion mask. Figure 5 shows the result of this experiment. Despite the fact that the MLP has no access to any iris texture information, the accuracy achieved is similar to that achieved on the iris images.

The results of this experiment suggest that there are two paths by which mascara can make it easier to identify female iris images. To the degree that eyelash occlusion is not well segmented, the eyelash occlusion that contaminates the iris texture will be darker with mascara than it is without. To the degree that mascara makes it easier to segment more of the eyelash occlusion, the masked area of the iris will be larger. By whichever path, when high gender-from-iris accuracy is found using a dataset in which many women wear mascara, it is difficult to know if the accuracy is truly due to gender-related patterns in the iris texture, or simply due to the presence of mascara.

7. Features: hand-crafted and data-driven

Approaches explored to extract discriminative features from the normalized iris images include hand-crafted features (e.g., Gabor filtering, LBP) and data-driven features, in which the raw pixel intensity is fed into neural networks that may "learn" features. All the experiments followed the same methodology, described in Section 3.

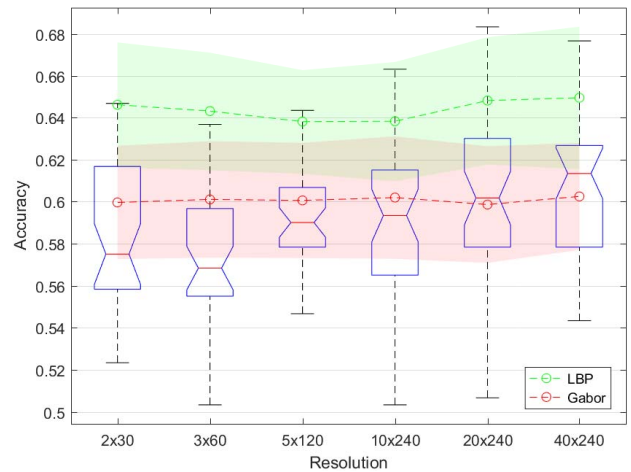


Figure 5: Accuracy using only the binary occlusion mask of the normalized iris. The dotted green line and shaded area denote the average accuracy and deviation for classifying normalized irises using LBP, and the dotted red line and shaded area are average accuracy and deviation using Gabor filtering. The box plots are accuracy distribution using simply the occlusion masks for the same normalized irises.

7.1. Data-Driven Features

Neural networks are an example of a classifier that can learn features from raw data. Data-driven features are "learned" from a dataset through a training procedure, and are dependent on the characteristics of the data and the classification goal. Here we present results of this approach, obtained through MLP and CNN classifiers. Details on the implementation of these networks are in Section 8.

Pixel intensity is the simplest feature. The pixel values of the masked, normalized image are fed directly to the neural network. Despite no explicit texture information being given to the network, the average accuracy of this approach was approximately 60%. This accuracy is similar to what

could be achieved using a simple intensity thresholding on the images, as seen in Section 5. This suggests that in this instance the neural network may be learning to predict gender based on a measure of average pixel intensity, or some other feature that is no more powerful.

Figure 6 shows a plot of the average accuracy obtained by this technique, across different image resolutions. It is worth observing that low resolutions like 2x30 and 3x60 the images could contain very little texture information because of the averaging of pixels.

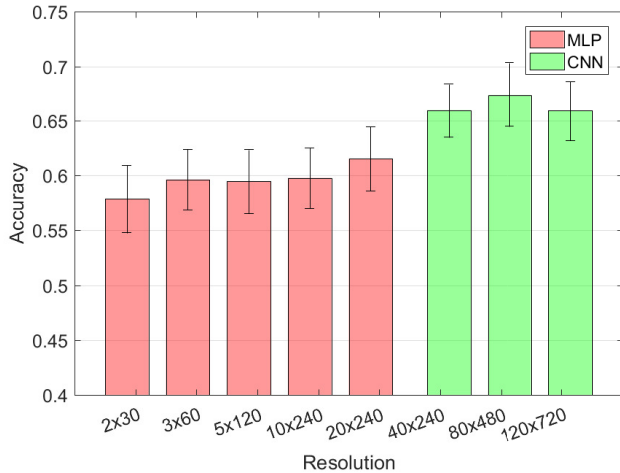


Figure 6: Accuracy based on pixel intensity for MLP and CNN classifiers. Average and standard deviation across 10 randomized train/test repetitions.

7.2. Hand-Crafted Features

Gabor filtering and LBP are popular examples of hand-crafted feature extraction techniques. Gabor filtering is done as part the standard approach to creating the "iris code" [3, 4]. In our experiments here, 1-Dimensional Gabor filtering was performed for each row of the normalized iris. We chose to explore a range of wavelengths similar to those used for iris recognition. LBP has been used in previous work on gender-from-iris [13].

For Gabor-filtered iris images, the average accuracy was $57\% \pm 3$ across all wavelengths, and there was no significant difference between different wavelengths considered. The fact that Gabor filtering resulted in worse classification than pixel intensity may seem surprising, but there is a possible explanation. Gabor filtering highlights the local occurrence of certain frequencies in the image by maximizing its response to these frequencies, while minimizing the response to other frequencies. If these low-response frequencies are related to features like occlusions or mascara, it makes sense that its attenuation has a negative effect on accuracy. As shown in section 5, the presence of eye cosmetics or even occlusion masks may artificially enhance the

gender classification accuracy, and their removal makes the problem harder. So these results are consistent with the idea that a significant part of the information that is used for gender classification may not come from the iris texture.

It is also important to mention that this work was limited to testing a certain range of parameters, based on those used for iris recognition. Since the main objective of iris recognition is to maximize the distinction between individual subjects and attenuate all non-person-specific features (such as gender, race, eye color, etc.), these parameters may not be the most appropriate for gender classification.

Local Binary Patterns (LBP) is a well-known method for texture analysis [10, 7]. We took some of the same LBP variations and parameters in [13], and used MLP neural networks to perform gender prediction.

In general, the best performances were achieved by uniform patterns and its variations (ULBP, CULBP-Mag and CULBP-Sign). ULBP histograms with and without patch overlapping had the highest accuracy with an average of 66%. Figure 7 shows an overall comparison between the three different feature extraction techniques. Gabor filtering had the worst results, with an average accuracy a little above 58%. In this graph, LBP extraction is divided into two different categories because of the significant performance difference between them. LBP images yields better accuracy than Gabor filtering or pixel intensity, but still well below Concatenated LBP Histograms.

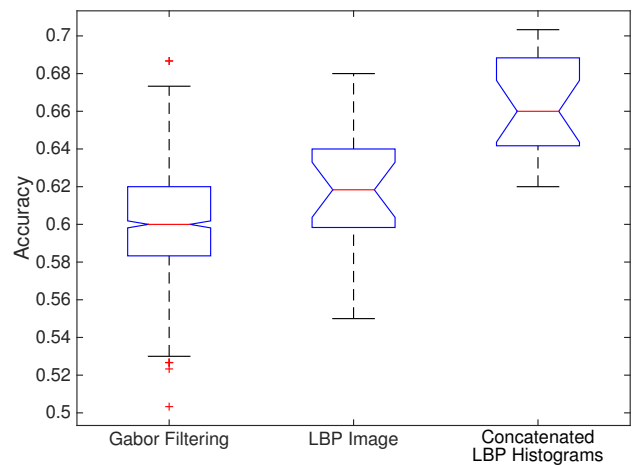


Figure 7: Classification accuracy distributions for hand-crafted features in a MLP classifier.

8. Neural Network Topologies

It is difficult to characterize the specific geometric or texture features that can be used to distinguish male from female irises. Thus, we decided to use an approach based on neural networks, so that they could learn the features that

are best fit for this classification.

The first portion of these experiments consists of an exploratory attempt to classify gender, training arbitrary-sized MLP Neural Networks using backpropagation. As a rule of thumb for the structuring of the networks, all of them had a first hidden layer of $1.5 \times P$, where P is the number of input features. The following layers of neurons in the network were defined as shown in Table 2. For example for a 10×240 image, the first network was configured as $3600 \times 20 \times 1$, the second $3600 \times 40 \times 1$, and so on.

In the cases where resolutions higher than 20×240 were used, the size of the MLP had to be reduced due to memory limitations. In these cases, we limited the size of Layer 1 to 5,000 neurons.

Layer 1	Layer 2	Layer 3	Layer 4	Output
$P \times 1.5$	20	-	-	1
$P \times 1.5$	40	-	-	1
$P \times 1.5$	300	40	-	1
$P \times 1.5$	300	80	-	1
$P \times 1.5$	600	80	-	1
$P \times 1.5$	300	80	20	1

Table 2: MLP neural network topologies used. P is the number of input pixels.

The activation function used for each layer of the network was a hyperbolic tangent, with the exception of the output layer, which consisted of a sigmoid activation function, in order to produce an output within the range of 0 and 1 corresponding to the gender.

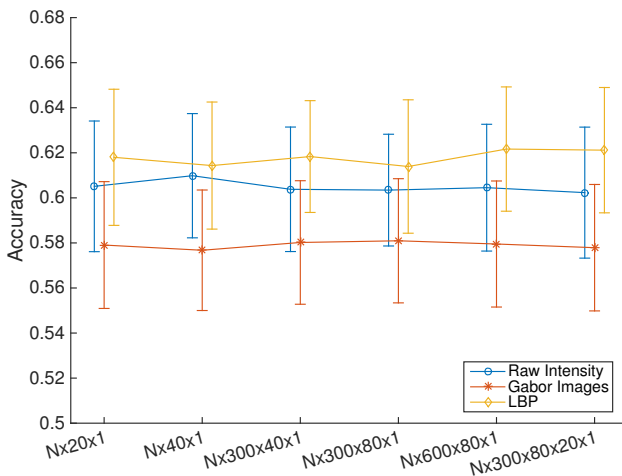


Figure 8: Accuracy of MLP classifier on Gender prediction, using different features according to network topology. Average and standard deviation across 10 randomized repetitions.

Network topology, within the range of options explored here, seems to have very little effect on the classification accuracy. Figure 8 shows how little variation occurs across different topologies with different types of image features. These results also emphasize that LBP features perform better than raw intensity or Gabor features.

8.1. Convolutional Neural Networks

We also experimented with a CNN architecture. These architectures have seen great progress in prominent image recognition benchmarks [11], [12], and their success in N-way image classification makes them promising for binary image classification as well. For the purposes of this paper, a CNN was used to classify gender based on two inputs: the full image and the segmented iris image with black occlusion masks (Figure 9, blue and red plots).

While the networks described in [11] and [12] are extremely large, trial by experimentation and difficulty of task (1000-way multi-scale classification vs 2-way single-scale classification) led to the conclusion that a smaller architecture would suffice in this environment. The network used consists of 3 sets of CNN layers, followed by 2 fully-connected (FC) layers and a softmax output. Each CNN set consisted of a Convolutional layer with a $(4, 4)$ kernel and a $(1, 1)$ stride, followed by a Max Pooling layer with a $(2, 2)$ kernel and a $(2, 2)$ stride. The number of features in each CNN layer were 16, 32 and 64 respectively, and the number of neurons in the FC layers were 1024 and 1536. Each neuron in the CNN and FC layers used the activation function $\max(0, X)$, commonly known as a Rectified Linear Unit, or ReLU activation.

Like before, GFI data was randomly split into 80/20 person-disjoint subsets for training/testing, and the network trained on 2500 batches of 32 images before testing in all cases. The training was carried out separately for left and right eyes on three different resolutions: 95×120 , 190×240 and 285×360 for the entire eyes, and 40×240 , 80×480 and 120×720 for normalized irises. Twenty randomized trials for each resolution and eye were performed.

Surprisingly, the results were virtually the same for all eyes and resolutions and almost identical to the accuracies obtained from using MLP networks. This may be because the data embedded in the image is low-level and separable by the MLP network, so the CNN layers simply transfer the underlying data to the final FC layers instead of extracting more information through its convolutions. This phenomenon would result in similar accuracies across network topologies and input resolutions, like those produced in this paper's experiments.

If we look at the resolutions used in the experiments with CNN and MLP on the entire eyes (Fig. 9, blue and black boxes), the lower resolution used with MLP shows there is no accuracy gain using larger images, or using a more com-

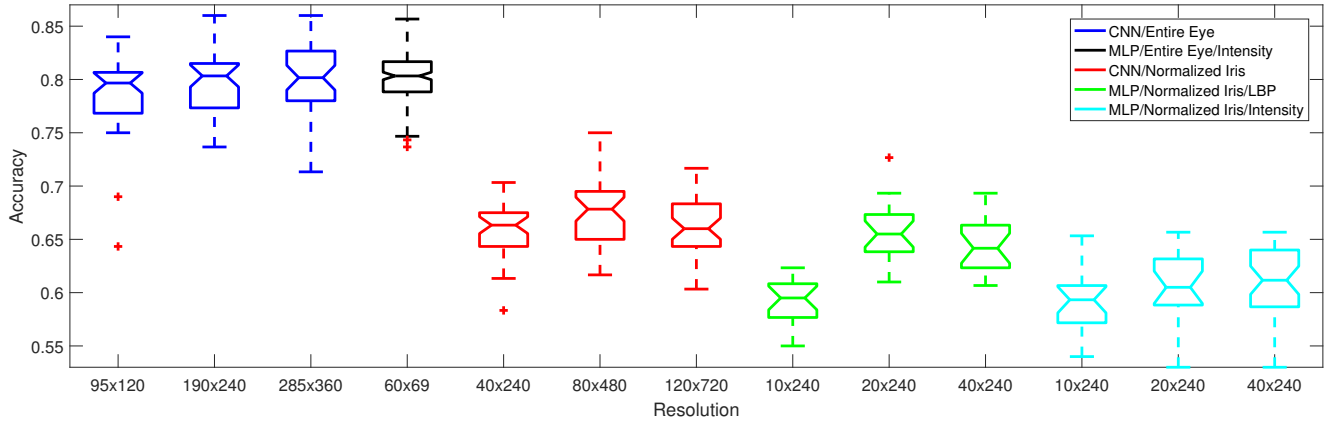


Figure 9: Comparison of CNN with MLP results. Blue and black boxes represent accuracy using the entire eye image, respectively for CNN and MLP. Red, green and cyan boxes denote the accuracy using the normalized iris image.

plex classifier. This means that classification is relying on image blobs that are large enough to be detected in a 60×69 image, once again suggesting that the fine details of iris texture do not contribute to gender classification as much as it was initially thought. When we look comparatively to normalized iris images (Fig. 9, red, green and cyan boxes), a significant portion of the gender-related information is lost. Again, CNNs do not seem to have a substantially higher accuracy.

9. Conclusions

We showed how the use of non-person-disjoint training and test can result in estimated gender-from-iris accuracy that is biased high. We also showed the importance of averaging over multiple trials. Using a single random train-test split of the data, the estimated accuracies ranged from 40% to 100%.

We showed that the presence of eye makeup results in higher estimated gender-from-iris accuracy. We also showed that classification based on the occlusion masks, disregarding completely the iris texture, results in an accuracy of approximately 60%. And we showed that simple averaging of the iris image intensity and thresholding can result in approximately 60% gender-from-iris accuracy.

Our experiments showed hand-crafted features like LBP to yield better prediction accuracy (66%) than data-driven features (60%) when using MLP networks. On the other hand, CNNs (using data-driven features) had performance comparable to MLPs+LBP. In a similar experiment using the entire eye images, CNNs and MLPs had equivalent performance (around 80%) using learned features.

Previous research may have misjudged the complexity of gender-from-iris, especially because of the subtle but important factors explored here. For future work, we suggest the creation of a subject-disjoint, mascara-free dataset. Cur-

rently, it is not clear what level of gender-from-iris accuracy is possible based solely on the iris texture.

10. Acknowledgements

The authors thank Dr. Adam Czajka for his invaluable insight and contribution.

This research was partially supported by the Brazilian Ministry of Education – CAPES through process BEX 12976/13-0.

References

- [1] A. Bansal, R. Agarwal, and R. K. Sharma. SVM based gender classification using iris images. *Proceedings - 4th International Conference on Computational Intelligence and Communication Networks, CICN 2012*, pages 425–429, 2012.
- [2] 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*, 11(3):441–467, 2016.
- [3] J. Daugman. How Iris Recognition Works. In *IEEE Transactions on Circuits and Systems for Video Technology*2, volume 14, pages 21–30. IEEE, 2004.
- [4] J. Daugman. Information Theory and the IrisCode. *IEEE Transactions on Information Forensics and Security*, 11(2):400–409, 2016.
- [5] J. S. Doyle, P. J. Flynn, and K. W. Bowyer. Effects of mascara on iris recognition. In I. Kakadiaris, W. J. Scheirer, and L. G. Hasebroock, editors, *Proc. SPIE 8712, Biometric and Surveillance Technology for Human and Activity Identification X*, volume 8712, page 87120L, may 2013.
- [6] M. Fairhurst, M. Erbilek, and M. D. Costa-Abreu. Exploring gender prediction from iris biometrics. In *Biometrics Special Interest Group (BIOSIG), 2015 International Conference of the*, pages 1–11, Sept 2015.

- [7] Z. Guo, L. Zhang, and D. Zhang. A completed modeling of local binary pattern operator for texture classification. *IEEE Transactions on Image Processing*, 19(6):1657–1663, 2010.
- [8] S. Lagree and K. W. Bowyer. Predicting ethnicity and gender from iris texture. In *IEEE International Conference on Technologies for Homeland Security (HST), 2011*, pages 440–445, Nov 2011.
- [9] X. Liu, K. W. Bowyer, and P. J. Flynn. Experiments with an Improved Iris Segmentation Algorithm. *Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05)*, (October):118–123, oct 2005.
- [10] 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, jul 2002.
- [11] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *3rd International Conference on Learning Representations (ICRL2015)*, May 2015.
- [12] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.
- [13] J. E. Tapia, C. A. Perez, and K. W. Bowyer. Gender Classification from Iris Images using Fusion of Uniform Local Binary Patterns. In *European Conference on Computer Vision (ECCV) Workshops, 2014*. Springer International Publishing, 2014.
- [14] J. E. Tapia, C. A. Perez, and K. W. Bowyer. Gender classification from the same iris code used for recognition. *IEEE Trans. Information Forensics and Security*, 11:1760–1770, 2016.
- [15] V. Thomas, N. V. Chawla, K. W. Bowyer, and P. J. Flynn. Learning to predict gender from iris images. In *First IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS), 2007.*, pages 1–5, Sept 2007.