

Predicting Ethnicity and Gender from Iris Texture

Stephen Lagree and Kevin W. Bowyer

Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, Indiana 46556 USA

stephen.lagree@gmail.com, kwb@cse.nd.edu

Abstract – Previous researchers have reported success in predicting ethnicity and in predicting gender from features of the iris texture. This paper is the first to consider both problems using similar experimental approaches. Contributions of this work include greater accuracy than previous work on predicting ethnicity from iris texture, empirical evidence that suggests that gender prediction is harder than ethnicity prediction, and empirical evidence that ethnicity prediction is more difficult for females than for males.

Keywords- *iris biometric; soft biometric; texture analysis; gender prediction; ethnicity prediction.*

I. INTRODUCTION

Iris biometrics derives an iris code from the texture pattern of the iris and uses this code to verify an identity claim with high accuracy, or to recognize a person as one from among a potentially large set of enrolled persons [1-5]. Iris biometrics is an active and expanding research area [3]. The technology has been used in border control in the United Arab Emirates for over a decade [4], and is currently used in large-scale applications such as India’s Aadhaar, or Unique ID, project to develop an identity system for 1.2 billion people [5].

This paper is concerned with analyzing iris texture in order to determine “soft biometric”, or demographic, attributes of a person, rather than identity. In particular, this paper is concerned with predicting the ethnicity and the gender of a person based on analysis of features of the iris texture.

The ability to predict soft biometric attributes such as ethnicity and gender based on iris texture is potentially useful in a variety of ways. In a 1-to-N recognition application, where a probe iris code is matched against a large number of enrolled identities for recognition, prediction of demographic factors could be used to order the search and thereby reduce the average search time. Also, a recognition system can naturally only recognize people who have previously been enrolled in the system. For persons who are not yet enrolled and so cannot be recognized, it may still be useful if the system could provide some basic demographic information for the person. In addition, at the time that a person is enrolled in the biometric system, prediction of demographic attributes could serve as an independent check on the accuracy of elements of information presented for enrollment. Also, there may be situations in which automated demographic classification and counting of persons without identity recognition is useful.

II. RELATED WORK

We are aware of just one previous publication on predicting gender from iris texture, by Thomas et al [6]. We are aware of just three previous publications dealing with predicting ethnicity from iris texture, two by Qiu et al [7,8] and an earlier version of our own work [16]. Work by Stark et al [20] is less directly related and looks at categorization of iris texture into general classes by human observers, and finds that some categories are correlated with ethnicity.

Thomas et al [6] experimented with iris images acquired using an LG 2200 iris sensor [9]. Like all commercial iris biometric sensors that we are aware of, these images are 640x480 in size, and acquired using near-infrared illumination. Using a version of the IrisBEE software from the Iris Challenge Evaluation [10], they segmented the iris region, created a normalized iris image, and created a log-Gabor filtered version of the normalized iris image. Their normalized iris image is 20 x 240 in size, and the complex-valued log-Gabor filtered version of the iris image can be viewed as having a 20x240 real part and a 20x240 imaginary part. They use only the real part to compute their texture features. They compute the mean value along each row of the real component of the log-Gabor-filtered image, for 20 features; the standard deviation along each row, for another 20 features; a measure of the local variance in a 1x5 window along each row, for another 20 features; and a measure of the local variation along each column computed over the whole image, for another feature. In addition to these texture features, they also use seven geometric features: (1) the difference in X between the pupil center and the iris center, (2) the difference in Y between the pupil center and the iris center, (3) the Euclidean distance between the pupil center and the iris center, (4) the area in pixels of the iris, (5) the area of the pupil, (6) the difference in area between the iris and the pupil, and (7) the ratio of the pupil area to the iris area. They report results of ten-fold cross-validation experiments with a C4.5 decision tree classifier, bagging with one hundred C4.5 decision trees and random subspaces of 5, 10, 15 and 20 features, with up to 100 decision trees. Using over 28,000 iris images, evenly split between males and females, and bagging with 100 decision trees, they were able to achieve gender prediction accuracy of nearly 75%. Their complete data set represented subjects of various ethnicities. In considering a

subset of data corresponding only to Caucasian subjects, they note that they were able to achieve over 80% accuracy. Their dataset represented images from about 300 different persons [21], and they used a person-disjoint partition for 10-fold cross-validation results, so that all images of a given person appeared in one fold and folds contained approximately equal numbers of images.

Qiu et al [7] developed an approach to classify iris textures into two ethnicity categories, Asian and non-Asian. Their 2,400 Asian images came from the CASIA dataset [11], and their 1,582 non-Asian images come from 384 UPOL images [12] and 1,198 UBIRIS images [13]. The 384 UPOL iris images are from 64 persons, 3 images of the left iris and 3 of the right iris. The UBIRIS images are from session one of the UBIRIS v1 dataset, representing approximately 240 different persons. The color images in the UBIRIS and UPOL datasets were converted to grayscale for this work. A training set of 1,200 images was created by randomly selecting 600 Asian and 600 non-Asian images, and the remaining 2,782 images were used as a test set. There does not appear to be any person-disjoint or iris-disjoint condition for the training and testing data in this experimental method; that is, it appears to allow for images from the same iris and same person to appear in both the training and the test data.

Qiu et al [7] construct a Gabor filter bank using four spatial orientations, six frequencies, and ten space constants, giving 240 Gabor filters. The filters are computed over two predetermined 30x256 regions of the normalized iris image, Region A that extends from the pupil out 3/8 of the way toward the iris boundary and Region B that extends the next 3/8 of the way out toward the iris boundary. The Gabor energy is computed for region B and the Gabor energy ratio is computed between the two regions. Using an AdaBoost approach to select six features from the set of 480 Gabor energy and Gabor energy ratio features, they were able to achieve 86.5% correct classification on the test set. They do not comment on the gender of the subjects represented in the image dataset.

In a later work, Qiu et al [8] take a different approach to computing texture features and use a different type of classifier. A filter bank of 40 even Gabor filters is used, formed by eight orientations and five scales. This gives a 40-dimensional feature vector for each pixel in the 60x256 ROI of each of 400 images. The 400 images are 200 from the CASIA dataset, representing Asian, and 200 from the BioSecure dataset, representing non-Asian. Using a K-means algorithm, 64 clusters are found, and the cluster centers are taken to represent “textons”; that is, commonly occurring fundamental texture elements. An image to be analyzed is then considered in terms of its 64-element texton histogram, where each pixel in the image is assigned to its closest texton. In this way, each image to be classified is represented by a 64-element feature vector. The experimental dataset comprises 2,400 images, formed as 20 images from each of 120 irises, representing 60 different persons, 30 from the CASIA dataset and 30 from the

BioSecure dataset [14]. A training set of 1,200 images is randomly selected, 600 from the CASIA dataset and 600 from the BioSecure dataset, and the remaining 1,200 images are the test set. Using a support vector machine classifier, they are able to achieve 88.3% correct classification on the test set. They also report 91.02% correct classification averaged between the training set and the test set. It appears that there is no person-disjoint or iris-disjoint condition enforced for the train and test data.

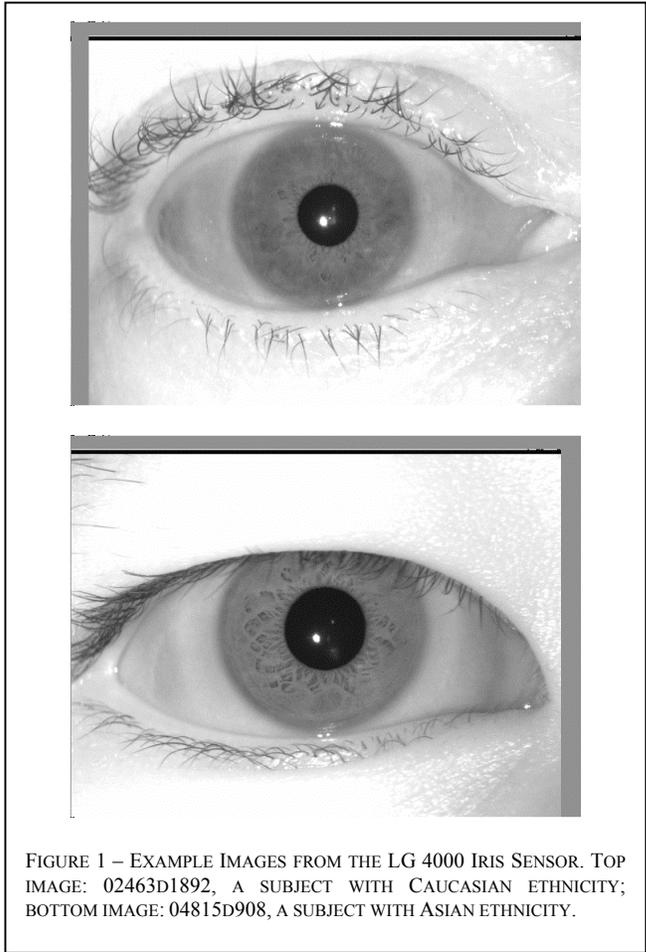


FIGURE 1 – EXAMPLE IMAGES FROM THE LG 4000 IRIS SENSOR. TOP IMAGE: 02463D1892, A SUBJECT WITH CAUCASIAN ETHNICITY; BOTTOM IMAGE: 04815D908, A SUBJECT WITH ASIAN ETHNICITY.

Our own earlier work in this area [16] addresses prediction of ethnicity based on iris texture. This current paper contains additional results related to ethnicity prediction and adds results related to gender prediction and to how one demographic factor affects the prediction of the other.

III. EXPERIMENTAL DATA

All images used in this work were obtained using an LG 4000 sensor [9]. The LG 4000 uses near-infrared illumination, and produces 480x640 images. Example images are shown in Figure 1. The IrisBee software [10] was used to segment the iris region in the image and create a 40x240 normalized iris image. The texture features are computed from this image. We do not use any information from the “ocular” region, such as the shape of the corner of the eye or the fold in the eyelid.

There is also a 40x240 mask that records the locations of occlusions from eyelid, eyelash, etc. The image segmentation and masking are exactly those that would be used by IrisBEE in processing the images for biometric recognition of a person’s identity. However, for the ethnicity and gender prediction results presented here, the normalized images are not processed by the log-Gabor filters that are used by IrisBEE to create the “iris code” for biometric purposes. (This is different from the features computed by Thomas et al [6].)

Texture features are computed by starting with various simple, basic texture filters. For a given point in the image, if applying a given filter would result in using any pixel that is masked as representing an occluded portion of the iris, then that filter application is skipped for that point. Six of the basic filters are “spot detectors” and “line detectors” of various sizes. Three other filters represent Laws’ texture measures S5S5, R5R5 and E5E5 [17]. Examples are depicted in Tables I to IX.

We compute texture features separately for eight five-pixel horizontal bands, running from the pupil-iris boundary out to the iris-sclera boundary, and ten twenty-four-pixel vertical bands of the 40x240 image. This is so that classification can be based on, for example, differences between the band of the iris nearest the pupil versus the band nearest the sclera.

We compute six summary statistics for each of the nine basic texture filters, for each of the eight horizontal regions of the image. The six statistics are: (1) average value of filter response over the region, (2) standard deviation of filter response, (3) 90th percentile value of filter response, (4) 10th percentile value of filter response, (5) range between 90th and 10th percentile value, and (6) a “local window difference” computed by subtracting the texture value at the current position from the average of the texture value at the two positions behind and the two positions ahead of it in the same row of the image. The six summary statistics x eight regions x nine texture features results in 432 features. We compute the first five summary statistics for each of the nine basic texture filters for each of the ten 24-pixel vertical regions. This results in another 450 features, for a total of 882 features. Compared to Thomas et al [6], we do not use any “geometric” features and we use a larger number of texture features.

TABLE I: SMALL “SPOT” DETECTOR

-1/8	-1/8	-1/8
-1/8	+1	-1/8
-1/8	-1/8	-1/8

TABLE II: LARGE “SPOT” DETECTOR

-1/16	-1/16	-1/16	-1/16	-1/16
-1/16	+1/9	+1/9	+1/9	-1/16
-1/16	+1/9	+1/9	+1/9	-1/16
-1/16	+1/9	+1/9	+1/9	-1/16
-1/16	-1/16	-1/16	-1/16	-1/16

TABLE III: VERTICAL LINE DETECTOR

-1/20	-1/20	+1/5	-1/20	-1/20
-1/20	-1/20	+1/5	-1/20	-1/20
-1/20	-1/20	+1/5	-1/20	-1/20
-1/20	-1/20	+1/5	-1/20	-1/20
-1/20	-1/20	+1/5	-1/20	-1/20

TABLE IV: HORIZONTAL LINE DETECTOR

-1/20	-1/20	-1/20	-1/20	-1/20
-1/20	-1/20	-1/20	-1/20	-1/20
+1/5	+1/5	+1/5	+1/5	+1/5
-1/20	-1/20	-1/20	-1/20	-1/20
-1/20	-1/20	-1/20	-1/20	-1/20

TABLE V: WIDE VERTICAL LINE DETECTOR

-1/10	+1/15	+1/15	+1/15	-1/10
-1/10	+1/15	+1/15	+1/15	-1/10
-1/10	+1/15	+1/15	+1/15	-1/10
-1/10	+1/15	+1/15	+1/15	-1/10
-1/10	+1/15	+1/15	+1/15	-1/10

TABLE VI: WIDE HORIZONTAL LINE DETECTOR

-1/10	-1/10	-1/10	-1/10	-1/10
+1/15	+1/15	+1/15	+1/15	+1/15
+1/15	+1/15	+1/15	+1/15	+1/15
+1/15	+1/15	+1/15	+1/15	+1/15
-1/10	-1/10	-1/10	-1/10	-1/10

TABLE VII: LAWS TEXTURE FILTER “S5S5”

+1	0	-2	0	1
0	0	0	0	0
-2	0	+4	0	-2
0	0	0	0	0
+1	0	-1	0	+1

TABLE VIII: LAWS TEXTURE FILTER “R5R5”

-1	-4	6	-4	+1
-4	+16	-24	+16	-4
6	-24	+36	-24	+6
-4	+16	-24	+16	-4
+1	-4	+6	-4	+1

TABLE IX: LAWS TEXTURE FILTER “E5E5”

1	2	0	-2	-1
2	4	0	-4	-2
0	0	0	0	0
-2	-4	0	4	2
-1	-2	0	2	1

IV. ETHNICITY PREDICTION

Our initial experiments were performed with a dataset of 1200 feature vectors, representing 60 Asian and 60 Caucasian

subjects. For each subject, we selected five images each of the left and the right iris. Multiple images of each iris are used in order to capture natural variation due to differences in occlusion, pupil dilation, lighting and other factors. We used **person-disjoint** 10-fold cross-validation, so that the dataset was randomly divided into 10 folds, with each fold having the images for 6 persons of each ethnicity. If the train and test data are person-disjoint, then the test data consists of images from people who have no image in the training data. Since it is known that there is a substantial degree of similarity in the texture pattern of the left and right iris of the same person [15], having person-disjoint train and test data should ensure that the performance on the test set is a better estimate of the ability to generalize to unseen persons.

We explored the performance that could be achieved in ethnicity classification using various classifiers available in Weka [18]. Results are summarized in Table VI. By varying parameters, we achieved performance gains on some of the classifiers. However, in this initial experiment we found the highest accuracy of 90.58% correct ethnicity prediction using the SMO support vector algorithm with Weka’s default parameter settings. While SMO gave the highest accuracy, the RandomForest and Bagged FT classifiers resulted in accuracy that was not significantly lower.

TABLE VI.: ETHNICITY PREDICTION RESULTS FOR VARIOUS CLASSIFIERS

Algorithm	Accuracy
SMO	90.58
Random Forest	89.50
Bagged FT	89.33
FT	87.67
ADTree	85.25
J48Graft	83.67
J48	83.08
Naïve Bayes	68.42

This performance on ethnicity prediction is an improvement over the 86.5% reported for the test set in [7] and the 88.3% reported for the test set in [8]. Additionally, the 90.58% result is obtained under the condition of person-disjoint train and test data. To judge the importance of this factor, we made an additional run of this experiment assigning images to folds randomly but without enforcing the person-disjoint condition. Randomizing across folds without enforcing a person-disjoint condition resulted in an accuracy of 96.17%, an estimate that is “optimistic” by over 5%.

To investigate the interaction of gender and ethnicity, we created four 600-image datasets: (1) 60 male subjects, 30 Asian and 30 Caucasian, with 5 images each of the left and right iris; (2) 60 female subjects, 30 Asian and 30 Caucasian; (3) 60 subjects split into 15 female Asian, 15 male Asian, 15 female Caucasian and 15 male Caucasian, formed by dividing the two datasets (1) and (2) into halves and combining the first halves of the two; and (4) another dataset of 60 subjects split into 15 female Asian, 15 male Asian, 15 female Caucasian and 15 male Caucasian, formed from the “other” halves of (1) and (2).

Based on the SMO classifier performance in our initial experiments, we trained SMO classifiers for ethnicity prediction on each of these four datasets. To explore the effects of varying the amount of training data, we trained using 2-fold, 5-fold and 10-fold person-disjoint cross-validation. For datasets (3) and (4), this resulted in every other partition in the 10-fold cross-validation alternating between 1 and 2 persons in a given gender-ethnicity combination.

Results for these classifiers are plotted in Figure 2. Results for datasets (3) and (4) are averaged and plotted as one curve labeled as representing “male and female” data in training. The accuracy of ethnicity prediction for this mixed-gender training data increases slightly from 2-fold to 5-fold to 10-fold cross-validation, as the amount of training data increases from 50% to 80% to 90% of the total data. Interestingly, there is a significant difference in the accuracy for ethnicity prediction in the single-gender scenarios. Ethnicity prediction for the case of all male subjects is above 90% for each of the 2-fold, 5-fold

FIGURE 2. CROSS-VALIDATION RESULTS OF ETHNICITY PREDICTION, OVERALL AND BY GENDER.

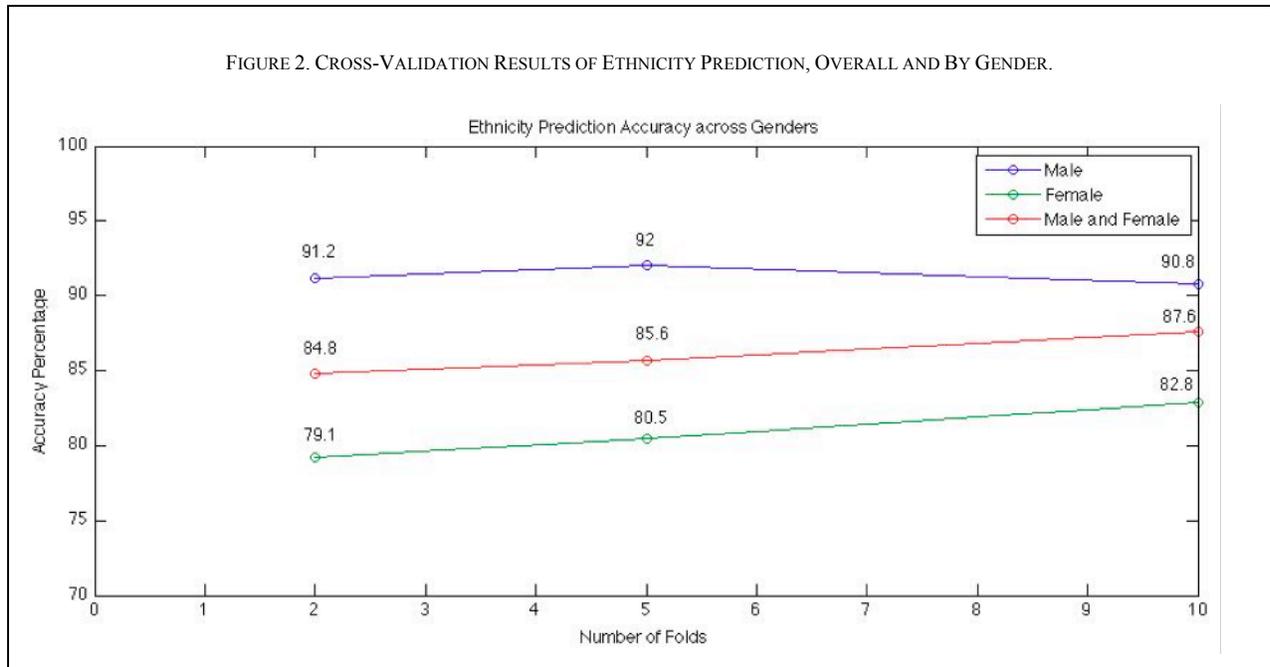
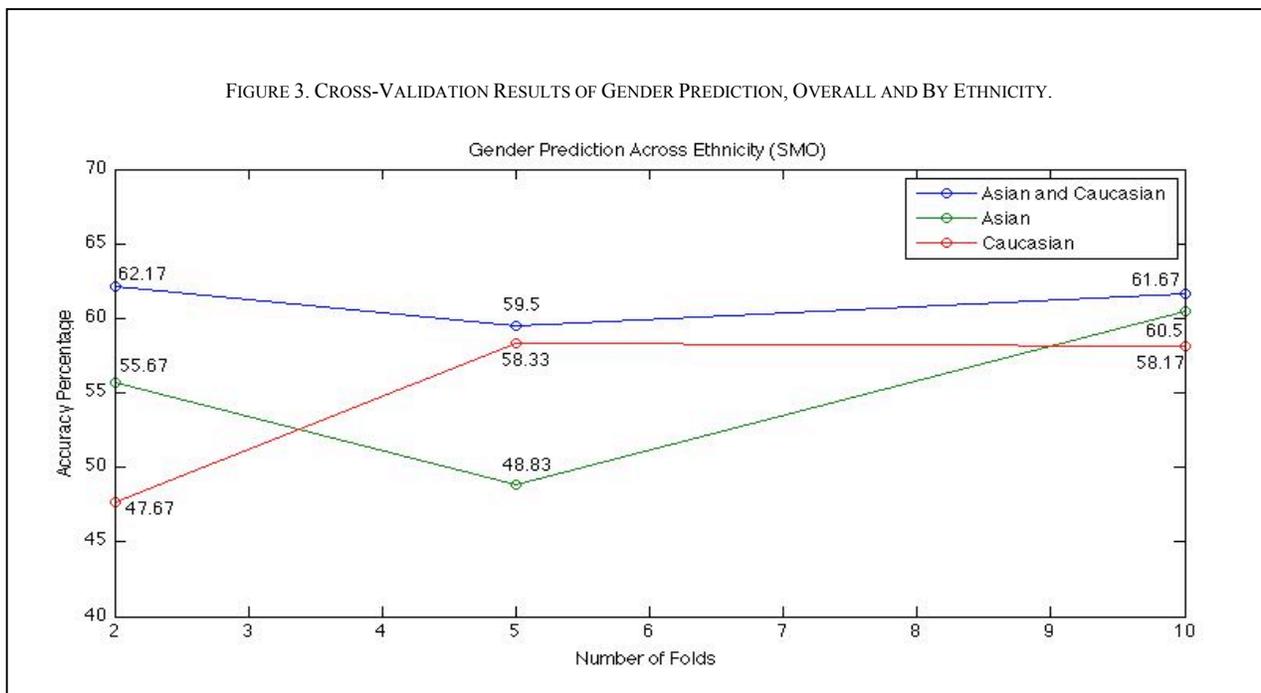


FIGURE 3. CROSS-VALIDATION RESULTS OF GENDER PREDICTION, OVERALL AND BY ETHNICITY.



and 10-fold results. However, ethnicity prediction starts out just below 80% and rises only to about 83% for the case of all female subjects.

Based on these results, it appears that predicting {Asian, Caucasian} ethnicity is more challenging for an all-female dataset than an all-male dataset. Training and testing a classifier on a mixed-gender dataset results in accuracy roughly the average of the two single-gender results. Also, it appears that the accuracy obtained in the single-gender case could be increased through the use of larger amounts of training data. The accuracy obtained for the mixed-gender datasets in this experiment using 600-image datasets is 87.6%, compared to 90.58% in the earlier experiment with the 1,200-image dataset. Interestingly, the results for the male-only and female-only datasets seem to indicate that increased training data size would benefit primarily the accuracy for females.

V. GENDER PREDICTION

Using the same four datasets as described in the previous section, we also trained SMO classifiers for gender prediction. The level of accuracy initially achieved for gender prediction seemed low relative to the accuracy that we achieved for ethnicity prediction and relative to the accuracy for gender prediction reported by Thomas et al [6]. Therefore, we explored several variations on the feature vector and classifier used for ethnicity prediction in the previous section. We also tried bagging on the data items used with the SMO classifier, generating 10 and 100 bags. However, neither of these resulted in any significant change in accuracy, and so they are not reported in the results summarized in Figure 3. Figure 3 summarizes the accuracy obtained using 2-fold, 5-fold and 10-fold cross-validation with the SMO classifier. Gender prediction on the mixed-ethnicity dataset reached only about 62% accuracy, and there was no significant difference

between this and the gender prediction accuracy for either single-ethnicity dataset.

The accuracy for gender prediction in Figure 3 is significantly lower than that reported in Thomas et al [6]. There are several possible explanations. One is a difference in dataset size. Experiments in [6] used over 28,000 images whereas our results in Figure 3 use a 600-image dataset, for about a factor of 50 difference in number of images in the training set. Another difference is in the feature vectors used. The results in [6] are obtained with features computed on the log-Gabor filtered version of the iris image, whereas we use features based on simple spot, line and Laws texture measures. A third difference is that the results in [6] were obtained with bagging 100 C4.5 decision trees, whereas our results were obtained with the Weka SMO support vector machine classifier.

VI. CONCLUSIONS

Previous research has considered either the problem of predicting ethnicity from iris features or the problem of predicting gender from iris features, but has not considered both problems using the same dataset and similar experimental approach. Our work is the first to report on predicting both ethnicity and gender, and on the mixed effects of the two problems. We find improved accuracy relative to previous work on predicting ethnicity from iris texture. Accuracy of predicting {Asian, Caucasian} ethnicity using person-disjoint 10-fold cross-validation on a 120-person, 1,200-image dataset exceeds 90%. Accuracy on predicting gender using person-disjoint 10-fold cross-validation on a 60-person, 600-image dataset is close to 62%. This is below the accuracy previously reported by Thomas et al [6] for gender prediction.

Based on our experimental results, it appears that predicting gender from iris is a more difficult problem than predicting {Asian, Caucasian} ethnicity. It also appears that predicting ethnicity is more difficult for females than for males. On the other hand, we do not see any evidence that the difficulty of predicting gender varies across ethnicity.

We explored the importance of the different basic texture features by computing results for the ethnicity prediction problem. The most important individual feature group was Laws E5E5 and the least important individual feature group was Laws R5R5.

TABLE IX. BASIC TEXTURE FILTERS WITH SMO ON ETHNICITY PREDICTION

Feature	Accuracy
Small Spot Detector	85.58
Large Spot Detector	85.67
Vertical Line Detector	87.42
Wide Vertical Line	85.50
Horizontal Line Detector	78.92
Wide Horizontal Line Detector	78.33
S5S5	78.17
R5R5	73.33
E5E5	88.0
All Features	90.58

We also performed a control experiment by computing results of a group A / B experiment with sixty subjects of data in each group and the groups balanced on gender and ethnicity. Thus in this experiment there is no overall difference in gender or ethnicity between the two groups. Performing a person-disjoint, 10-fold cross-validation with the SMO, FT and J48 classifiers results in 48.4%, 50.3% and 50.5% accuracy, respectively, in predicting group A / group B. This is not significantly different from the expected 50%, indicating that our experiments focused on ethnicity or gender are not affected by a hidden flaw in the methodology.

Future work should include exploration of additional classifiers and features, especially for use in gender prediction. For example, because it is known that iris texture appearance may vary with changes in pupil dilation [19], adding the pupil dilation ratio as a feature may improve performance.

ACKNOWLEDGMENT

This research was funded by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), through the Army Research Laboratory (ARL). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing official policies, either expressed or implied, of IARPA, the ODNI, the Army Research Laboratory, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints

for Government purposes notwithstanding any copyright notation herein. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of our sponsors.

Stephen Lagree was supported by the Ateyeh Endowment for Excellence at the University of Notre Dame as an Ateyeh Undergraduate Research Scholar.

REFERENCES

- [1] J. Daugman, "Probing the uniqueness and randomness of iris codes: results from 200 billion iris pair comparisons", *Proceedings of the IEEE*, Nov. 2006, 94 (11), 1927 – 1935.
- [2] J. Daugman, "How iris recognition works", *IEEE Transactions on Circuits and Systems for Video Technology* 14 (1), January 2004, 21-30.
- [3] K. W. Bowyer, K. Hollingsworth and P. J. Flynn, "Image understanding for iris biometrics: a survey", *Computer Vision and Image Understanding* 110(2), 281-307, May 2008.
- [4] J. Daugman and I. Malhas, "Iris recognition border crossing system in the UAE", *International Airport Review* (2), 2004.
- [5] J. J. Romero, "Fast start for world's biggest biometrics ID project", *IEEE Spectrum*, May 2011.
- [6] V. Thomas, N. Chawla, K.W. Bowyer and P.J. Flynn, "Learning to predict gender from iris images", *IEEE Int. Conf. on Biometrics: Theory, Applications, and Systems (BTAS)*, Sept 2007.
- [7] X. C. Qiu, Z. A. Sun, and T.N. Tan, "Global texture analysis of iris images for ethnic classification", *Springer LNCS 3832: Int. Conf. on Biometrics*, pages 411-418, June 2006.
- [8] X. C. Qiu, Z. A. Sun, and T. N. Tan, "Learning appearance primitives of iris images for ethnic classification", *IEEE Int. Conf. on Image Processing (ICIP)*, pages II: 405-408, 2007.
- [9] LG. <http://www.lgiris.com/>
- [10] P. J. Phillips, W. T. Scruggs, A. O'Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott and M. Sharpe, "FRVT 2006 and ICE 2006 Large-Scale Experimental Results", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32 (5), May 2010, 831-846.
- [11] CASIA Iris Database, <http://www.cbsr.ia.ac.cn/english/IrisDatabase.asp>
- [12] UPOL Iris Database, <http://phoenix.inf.upol.cz/iris/>
- [13] UBIRIS Iris Database, <http://iris.di.ubi.pt/>
- [14] Biosecure multimodal datasets, <http://biosecure.it-sudparis.eu/AB/>
- [15] K.W. Bowyer, S. Lagree and S. Fenker, "Human versus biometric detection of similarity in left and right irises", *IEEE International Carnahan Conference on Security*, October 2010.
- [16] S. Lagree and K.W. Bowyer, "Ethnicity prediction based on iris texture features", *22nd Midwest Artificial Intelligence and Cognitive Science Conference (MAICS)*, April 2010.
- [17] K. Laws, *Textured Image Segmentation*, Ph.D. Dissertation, University of Southern California, January 1980.
- [18] Weka 3. <http://www.cs.waikato.ac.nz/ml/weka/>
- [19] K. Hollingsworth, K.W. Bowyer and P.J. Flynn, "Pupil dilation degrades iris biometric performance", *Computer Vision and Image Understanding* 113 (1), January 2009, 150-157.
- [20] L. Stark, K.W. Bowyer and S. Siena, "Human perceptual categorization of iris texture patterns", *IEEE Int. Conf. on Biometrics: Theory, Applications, and Systems (BTAS)*, Sept 2010.
- [21] K.W. Bowyer and P.J. Flynn, "The ND-IRIS-0405 iris image dataset", <http://www.nd.edu/~cvrl/papers/ND-Iris-0405.pdf>