Ethnicity Prediction Based on Iris Texture Features

Stephen Lagree and Kevin W. Bowyer
Department of Computer Science and Engineering
University of Notre Dame
Notre Dame, Indiana 46556 USA
slagree@nd.edu, kwb@cse.nd.edu

Abstract

This paper examines the possibility of predicting ethnicity based on iris texture. This is possible if there are similarities of the iris texture of a certain ethnicity, and these similarities differ from ethnicity to ethnicity. This sort of "soft biometric" prediction could be used, for example, to narrow the search of an enrollment database for a match to probe sample. Using an iris image dataset representing 120 persons and 10-fold person-disjoint cross validation, we obtain 91% correct Asian / Caucasian ethnicity classification.

Introduction

Iris texture has been shown to be useful for biometric identification and verification (Bowyer, Hollingsworth, and Flynn 2008; Phillips et al. 2005; Phillips et al. 2010; Daugman 2006). Studies have been done to determine if iris texture contains information that can determine "soft biometric" attributes of a person, such as ethnicity (Qiu, Sun, and Tan 2006; Qiu, Sun, and Tan 2007a) or gender (Thomas et al. 2007). This paper analyzes the possibility of ethnicity prediction based on iris texture. The ability of biometric systems to recognize the ethnicity of a subject could allow automatic classification without human input. Also, in an iris recognition system, an identification request includes a "probe" iris, which is checked against a "gallery" of enrolled images, to find the correct identity of the requested iris. One application of this feature is to narrow down the gallery of subjects to compare an iris to for identification purposes. In a system with millions of enrolled subjects, comparing an iris to all subjects could take an extremely long time. Narrowing down the gallery to only irises with the same ethnicity as the probe iris for comparison could give a great speed improvement.

Related Work

The CASIA biometrics research group has performed research on iris texture elements, including studies (Qiu, Sun, and Tan 2006; Qiu, Sun, and Tan 2007a; Qiu, Sun, and Tan 2007b) on determining ethnicity based on iris texture. To our knowledge, this is the only other work on predicting ethnicity from iris texture. In (Qiu, Sun, and Tan 2006), they report 86% accuracy in Asian / Caucasian classification. Thomas et al. (2007) suggests that the work in (Qiu, Sun, and Tan 2006) may be biased due to illumination differences in the two datasets the images were taken from, the Asian subject images coming from one dataset and the Caucasian subject images from another dataset. If one dataset was generally brighter or darker than the other, this factor could have entered into the learned algorithm for separating the subjects based on lighting, not iris texture. In the results presented in this paper, we eliminate this issue by using images taken from a single database to build our classifier, so that any acquisition setup differences are just as likely to appear in either ethnicity class. In (Qiu, Sun, and Tan 2007a), the CASIA group reports 91% accuracy in Asian / non-Asian ethnicity classification, using support vector machines and texton features. The dataset in this work is composed of 2,400 images representing 60 different persons, so that there are 20 images per iris. They divide the dataset into a 1,200-image training set and a 1,200-image test set, with training and test set not specified to be person-disjoint. In general, if iris images from the same person appear in both the training and the test set, then the performance estimate obtained is optimistically biased. In the results presented in this paper, we eliminate this issue by using a person-disjoint ten-fold cross-validation.

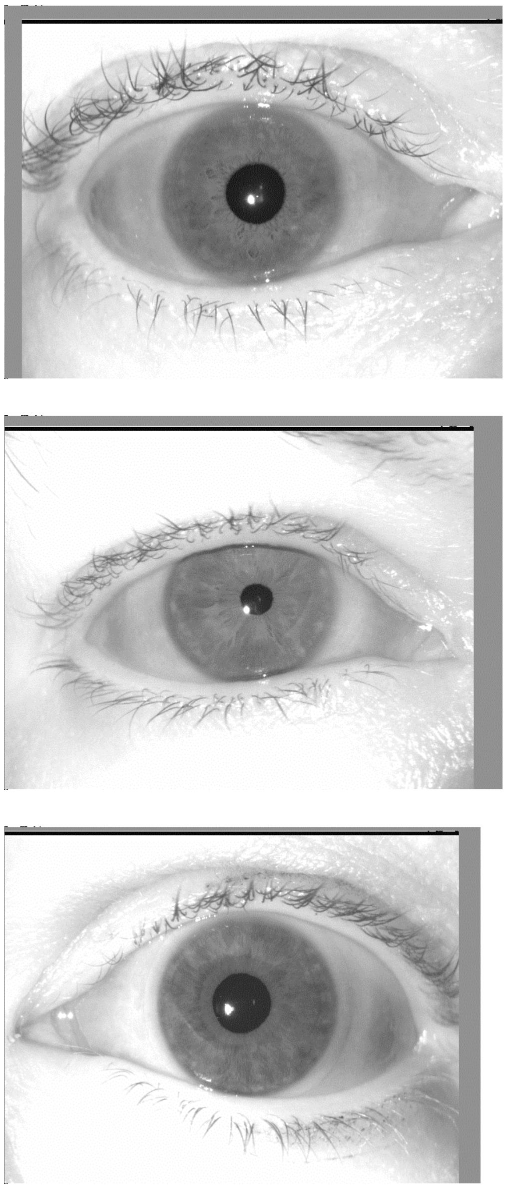

Figure 1 – Example LG 4000 Iris Images From Subjects with Caucasian Ethnicity (top: image 02463d1892; middle: image 04327d1264; bottom: image 04397d1461).

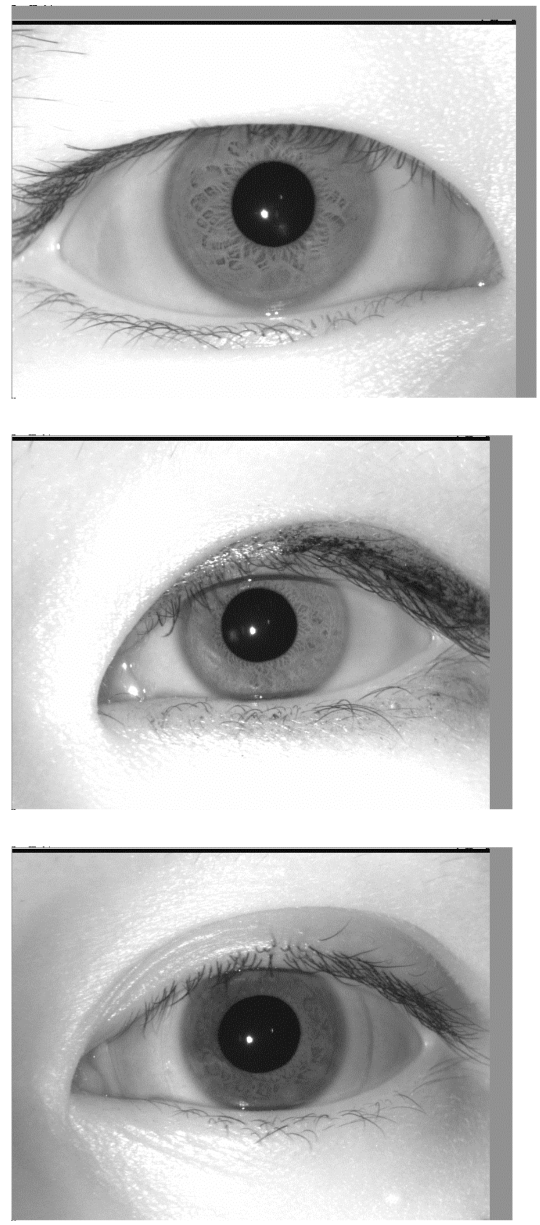

Figure 2 – Example LG 4000 Iris Images From Subjects with Asian Ethnicity (top: image 04815d908; middle: image 04629d1385; bottom: image 05404d80).

In a study of how human observers categorize images, Stark, Bowyer, and Siena (2010) found that humans perceive general differences in iris texture that can be used to classify iris textures into categories of similar texture pattern. Observers grouped a set of 100 iris images into categories of similar texture. The 100 images represented 100 different persons, and the 100 persons were balanced on gender and on Asian / Caucasian ethnicity. The observers did not know the gender or ethnicity of the persons in the iris images. However, the grouping of images into categories of similar iris texture resulted in categories that were, on average, split 80% / 20% on ethnicity. The same categories were on average divided much more closely to 50% / 50% on gender. Thus, one result of Stark's work (2010) is that human observers perceive consistent ethnicity-related differences in iris texture. In this paper, we want to train a classifier to explicitly perform the sort of ethnicity classification that was found as a side effect of the texture similarity grouping done by humans in (Stark, Bowyer, and Siena 2010) and that was previously explored in (Qiu, Sun, and Tan 2006; Qiu, Sun, and Tan 2007a).

Dataset

We want to see how accurately we can identify ethnicity based on iris texture. For this study we will use two ethnicity classes, Caucasian and Asian. This study used 1200 iris images selected from the University of Notre Dame's iris image database. (This is a newer database than was released to the iris biometrics research community for the government's Iris Challenge Evaluation (ICE) program (Phillips et al. 2005; Phillips et al. 2010).) All images were obtained using an LG 4000 sensor at Notre Dame. As with all commercial iris biometrics systems that we are aware of, the images are obtained using near-infrared illumination, and are 480x640 in size. One half of the images, 600, were of persons whose ethnicity is classified as Asian and the other half were from persons classified as Caucasian. For each ethnicity, the 600 images represented 60 different persons, with 5 left iris images and 5 right iris images per person. This 1,200-image dataset was randomly divided into 10 folds of 120 images each, with 6 persons of each ethnicity in each fold. Thus the images in the folds are person-disjoint; that is, each person's images appear in just one fold.

Segmentation

For this iris texture prediction study, we want to base our findings solely on iris texture. Therefore we exclude periocular clues that might be used as an indicator of ethnicity. We segment the images to obtain the region of interest, and mask out the eyelid-occluded portions of the iris. We use Notre Dame's IrisBee software to perform the segmentations (Phillips et al. 2005). The output from IrisBee that we use for texture examination is a 240x40 pixel normalized iris image along with the corresponding bitmask of eyelid and eyelash occlusion locations. 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, the normalized images are not processed by the log-Gabor filters that are used by IrisBEE to create the "iris code" for biometric recognition. We create a different texture feature vector for ethnicity prediction.

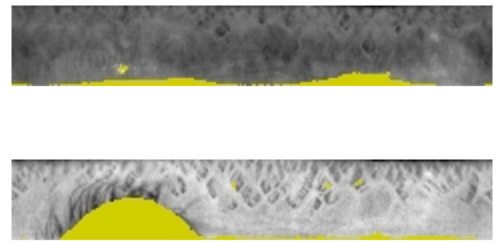

Figure 3 – Examples of Segmented, Normalized Iris Images. (top: normalized image derived from image 02463d1892 above; bottom: normalized image derived from image 04815d908 above). The green regions indicate the "mask" of where the iris texture

Feature Generation

After an image is segmented and normalized, we compute texture features that can be used in training a classifier to categorize images according to ethnicity. To do this we apply different filters to the image at every non-masked pixel location, and use the results of the filter to build a feature vector. Six of the filters we have used are "spot detectors" and "line detectors" of various sizes, as depicted in Tables I to VI. For a given point in the image, if applying a given filter would result in using any pixel that is masked, then that filter application is skipped for that point. The rest of the filters, depicted in Tables VI-VIII, were created using Laws' Texture Measures (Laws 1980). These are designed to give responses for various types of textures when convolved with images.

A feature vector that describes the texture is computed for each iris image. We divided the normalized image array into a number of smaller sections in order to compute statistics for sub-regions of the normalized image. This is so that classification could be based on, for example, relative differences between the band of the iris nearer the pupil versus the band of the iris furthest from the pupil. These regions were ten four-pixel horizontal bands and four 60-pixel vertical bands of neighboring pixels in the normalized iris image. The ten horizontal bands correspond to concentric circular bands of the iris, running from the pupil out to the sclera (white) of the eye. The four vertical bands correspond roughly to the top, right, bottom and left parts of the iris. Since the filters are looking for different phenomena in the image, we find statistics for the filter response of each image. Each image contains 630 features, with 5 statistics calculated for each of the 9 filters on all of the 14 regions. The five statistics are: (1) average value of filter response, (2) standard deviation of filter response, (3) 90th percentile value of filter response, (4) 10th percentile value of filter response, and (5) range between 90th and 10th percentile value. The motivation for using the average value is to represent the strength of a given spot size or line width in the texture. The motivation for using the standard deviation is to represent the degree of variation in the response. The motivation for using the percentiles and range is to have an alternate representation of the variation that is not affected by small amounts of image segmentation error.

TABLE I: Small Spot Detector Filter

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

TABLE II: Large Spot Detector Filter

-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 Filter

-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: Wide Vertical Line Detector Filter

-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 V: Horizontal Line Detector Filter

-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 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: 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: 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

Results

We tried a variety of different classification algorithms included in the WEKA package (Weka). This included using meta-algorithms like Bagging with other classifiers. By changing parameters, we achieved performance gains on some of the algorithms. However, we found our best results using the SMO algorithm with the default parameters in WEKA for classification. The SMO algorithm implements "Sequential Minimal Optimization", John Platt's algorithm for building a support vector machine classifier (Weka). The input to the SMO algorithm is the feature vectors of all 1200 iris images that we have computed. To assess the results of our classifier we use cross-fold validation with ten folds using stratification based on ethnicity. These folds are also subject-disjoint to ensure the persons whose images are in the test data have not been seen by the classification algorithm in the training data.

The SMO classifier results in higher accuracy compared to a broad range of other classifiers, including decision tree based algorithms and bagging. Using Bagging on the top two classifiers, SMO and Random Forest, did not improve performance. Running the experiment with the SMO classifier and the feature vector as described above gives us an accuracy of 90.58%. This is good accuracy, representing an improvement on the 86% reported in (Qiu, Sun, and Tan 2006) and close to the 91% reported in (Qiu, Sun, and Tan 2007a) for a train-test split that was not person-disjoint. When we do not use person disjoint results, we see an accuracy of 96.17%, which is significantly higher than Qiu, Sun, and Tan (2006; 2007a) reported.

We computed the classification accuracy for each feature separately to see the impact of individual features. Table X shows that some of the single features have almost have the performance of all of the features together. However none of them do as well as the combination of all of the features. Some filters may be redundant; a combination of a few might reproduce the performance of all nine filters.

To ensure that the size of our training dataset was not limiting our accuracy levels, we ran the classifier with different numbers of folds. Table XI shows the results we achieved using 5, 10, and 20 fold cross validation. The accuracy levels are all within one percent, indicating that our performance should not be limited by our dataset size.

TABLE IX: Results for Different Classifiers

Algorithm	**Accuracy (%)**
SMO	90.58
RandomForest (100 Trees/Features)	89.50
Bagged FT	89.33
FT	87.67
ADTree	85.25
J48Graft	83.67
J48	83.08
Naïve Bayes	68.42

TABLE X: Feature Performance with SMO

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

TABLE XI: SMO By Number of Folds Used in Cross Validation

Folds	**Accuracy (%)**
5	90.00
10	90.583
20	90.1667

TABLE XII: SMO Accuracy By Fold Using 10 Fold Cross Validation

Fold	**Accuracy (%)**
1	91.667
2	100.000
3	88.333
4	90.833
5	97.500
6	82.500
7	98.333
8	90.000
9	87.500
10	79.167
Average	**90.583**

Future Work

To achieve even greater accuracy, we intend to implement additional and more sophisticated features, and to look at the effects of the size of the training set. We envision that the number of different persons represented in the training data is likely to be more important than the number of images in the training set; that is, doubling the training set by using twice as many images per person is likely not as powerful as doubling the number of persons.

For this experiment, we only looked at very broad ethnicity classifications. More work could be done to examine finer categories, such as Indian and Southeast Asian. The performance of a classifier such as this has not been tested on subjects of multiple ethnic backgrounds either.

Acknowledgments

This work is supported by the Technical Support Working Group under US Army contract W91CRB-08-C-0093, and by the Central Intelligence Agency. 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.

References

Bowyer, K.W.; Hollingsworth, K.; and Flynn, P. J. Image Understanding for Iris Biometrics: A Survey, *Computer Vision and Image Understanding,* 110(2), 281-307, May 2008.

Daugman, J., 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.

Laws, K. Textured Image Segmentation, Ph.D. Dissertation, University of Southern California, January 1980.

Phillips, P. J.; Bowyer, K. W.; Flynn, P.J; Liu, X; and Scruggs, T. W. The Iris Challenge Evaluation 2005, *Biometrics: Theory, Applications and Systems (BTAS 08)*, September 2008, Washington, DC.

Phillips, P. J.; Scruggs, W. T.; O'Toole, A.; Flynn, P.J.; Bowyer, K.W.; Schott, C. L.; and Sharpe, M. FRVT 2006 and ICE 2006 Large-Scale Experimental Results, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32 (5), May 2010, 831-846..

Stark, L; Bowyer, K.W.; and Siena,S. Human perceptual categorization of iris texture patterns, *Biometrics Theory, Applications and Systems (BTAS),* September 2010.

Thomas, V; Chawla, N; Bowyer, K. W.; and Flynn, P. J. Learning to predict gender from iris images. In *Proc. IEEE Int. Conf. on Biometrics: Theory, Applications, and Systems,* Sept 2007.

Qiu, X. C.; Sun, Z. A.; and Tan, T. N. Global texture analysis of iris images for ethnic classification. In *Springer LNCS 3832: Int. Conf. on Biometrics*, pages 411-418, June 2006.

Qiu, X. C.; Sun, Z. A.; and Tan, T. N. Learning appearance primitives of iris images for ethnic classification. In Int. Conf. on Image Processing, pages II: 405–408, 2007a.

Qiu, X. C.; Sun, Z. A.; and Tan, T. N. Coarse iris classification by learned visual dictionary. In Springer LNCS 4642: Int. Conf. on Biometrics, pages 770–779, Aug 2007b.

Weka 3. http://www.cs.waikato.ac.nz/ml/weka/.