The Prediction of Old and Young Subjects from Iris Texture

Amanda Sgroi, Kevin W. Bowyer, and Patrick J. Flynn
University of Notre Dame
asgroi kwb flynn @nd.edu

Abstract

Researchers have previously studied the prediction of "soft biometric" attributes such as gender and ethnicity from iris texture images. We present the results of an initial study to predict the relative age of a person from such images. We conclude that is possible to categorize iris images as representing a young or older person at levels of accuracy statistically significantly greater than random chance. This suggests that there may in fact be age-related information available in the iris texture, and motivates further study of this topic.

1. Introduction

The aging of the iris, and subsequent changes in the iris texture, is a topic of current interest in iris biometrics. Many researchers have reported on experiments that involve images of the same iris taken over a period of several years, and found that an iris template aging effect exists [1][9][3]. The template aging effect takes the form of an increase in the false nonmatch rate with increased time. Others have looked into the physical effects of aging on the iris and the consequences of these changes on iris biometrics[7][8][17]. Yet, there is no current work on the prediction of age from iris texture.

There are several works on face age estimation [10]. Classifiers such as the nearest neighbor classifier can be trained to predict the age of a subject [5]. Age estimation can also be modeled as a regression problem[11][13][12]. The optimal parameters for a regression function are learned from a set of training images, and the regression problem is solved for the age of a presented testing image. A combination of classification and regression approaches may also be used.

The aging of the human face is familiar and easily identified. The aging of the human eye is less visually obvious. The pace at which the human eye ages is unknown, but several factors such as average pupil size [2][20] and changes in corneal shape [15][6] have been documented to occur with aging. Through the experiments presented in this work, we explore a classification technique which seeks to categorize a person as younger or older using features from their iris's texture. The texture features and classifier ensembles used are similar to those used by previous researchers for the prediction of gender and ethnicity from iris texture images.

In this paper, we first present an overview of the data set, feature selection, and techniques used for developing the classifier. The baseline result is then presented with additional results on single feature performance. Lastly, conclusions and a discussion of potential future experiments are provided.

2. Data set

In this experiment, 50 subjects between the ages of 22 and 25 were selected as the "younger" group, and 50 subjects older than the age of 35 were selected as the "older" group. Six images were selected for each subject, 3 left eye images and 3 right eye images. All images were taken with the LG IrisAccess 4000 across various days at different indoor locations at the University of Notre Dame. Additionally, each pair of left and right eye images for a given person were acquired on different days. However, two exceptions to this condition were required in the old subject data set. Namely, two of the subjects only attended one acquisition session and thus only four images, 2 left eyes and 2 right eyes from the same day were included in this set. These subjects could not be replaced due to the scarcity of subjects older than 35 in our pool of data. Given these restrictions a total of 596 images were used in this study, 300 for the younger group and 296 for the older group.

The old and young data sets are also balanced with respect to gender, race, and eye color. Each set contains 18 male and 32 female subjects. Five races are represented in each set, which were self-reported by each subject. There are 2 Asian subjects, 1 Middle Eastern subject, 1 African American subject, 44 Caucasian subjects, and 1 subject of unknown race in each of the younger and older groups. Six eye colors are also represented and were self reported by the subjects. There are 11 subjects with blue eyes, 21 with brown eyes, 9 with green eyes, 7 with hazel eyes, 1 with

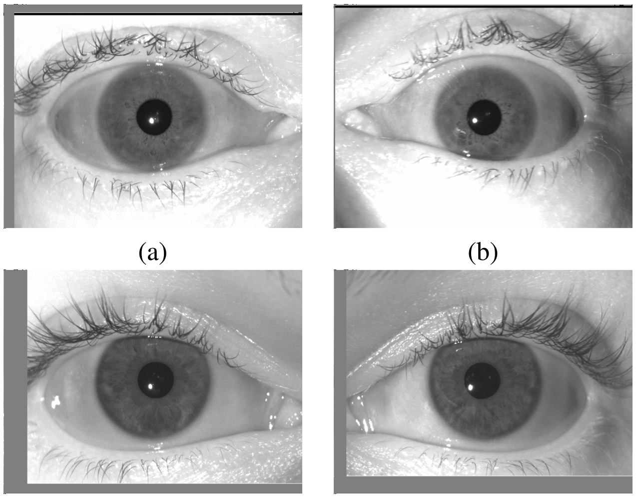


Figure 1. Right (a) and Left (b) eyes from old subject nd1S02463. Right (c) and Left (d) eyes from young subject nd1S06000.

black eyes, and 1 with gray eyes. Although the distribution of each covariate (gender, race, and eye color) is balanced, the subjects are not balanced along all three covariates. Figure 1 shows sample images from a subject in the old data set and a subject in the young data set. The age category to assign to these irises is not obvious to an observer.

As an initial exploration of the data set, an all-vs-all comparison within the old and young sets was performed using the IrisBEE matcher. Figure 3 depicts the score distribution and ROC curve results. Although the match and nonmatch score distributions look similar, the ROC curves reveal that young subjects are classified more accurately than images of the old subjects. This difference is small and may not be statistically significant.

3. Feature Extraction

To classify these data sets, features were computed from the segmented and normalized iris texture. Each image was first segmented using the IrisBEE preprocessor [19] and checked for quality. The unrolled irises were then used to create 630 features based on nine different filter responses in particular regions of the image defined by neighboring rows and columns. For each filter response, 70 features were generated. Figure 2 presents a sample segmented, normalized, and unrolled iris texture image, the nine resulting filter responses, as well as a description of the regions used per feature calculation. Table 1 further describes each filters' defining characteristic.

To extract our feature values, the filtered unrolled iris image is broken up into four row regions each containing 10 rows of size 240-by-10, and 10 column regions each containing 24 columns of size 24-by-40. The segmentation of

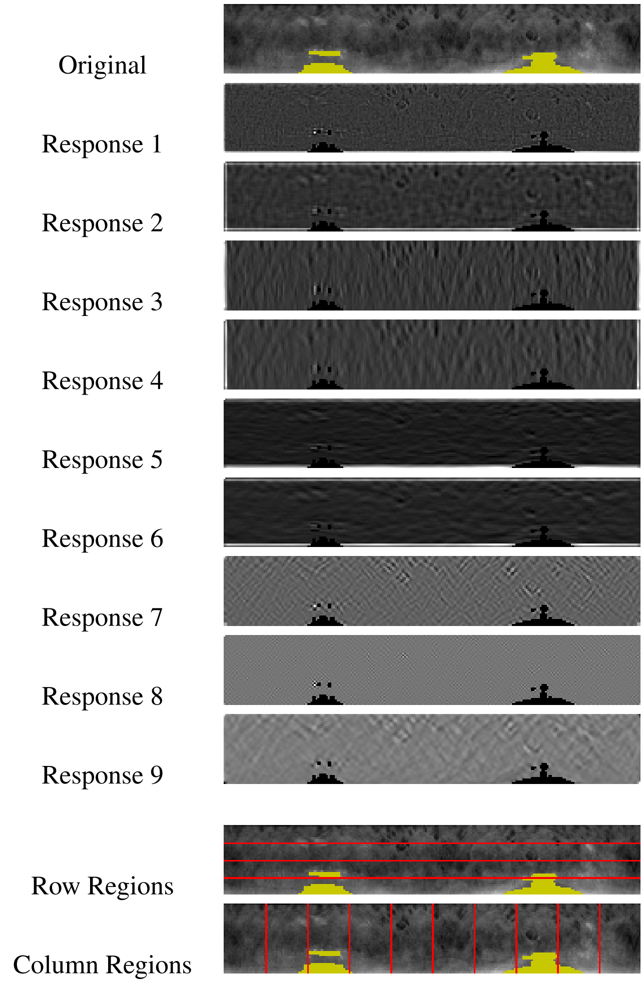
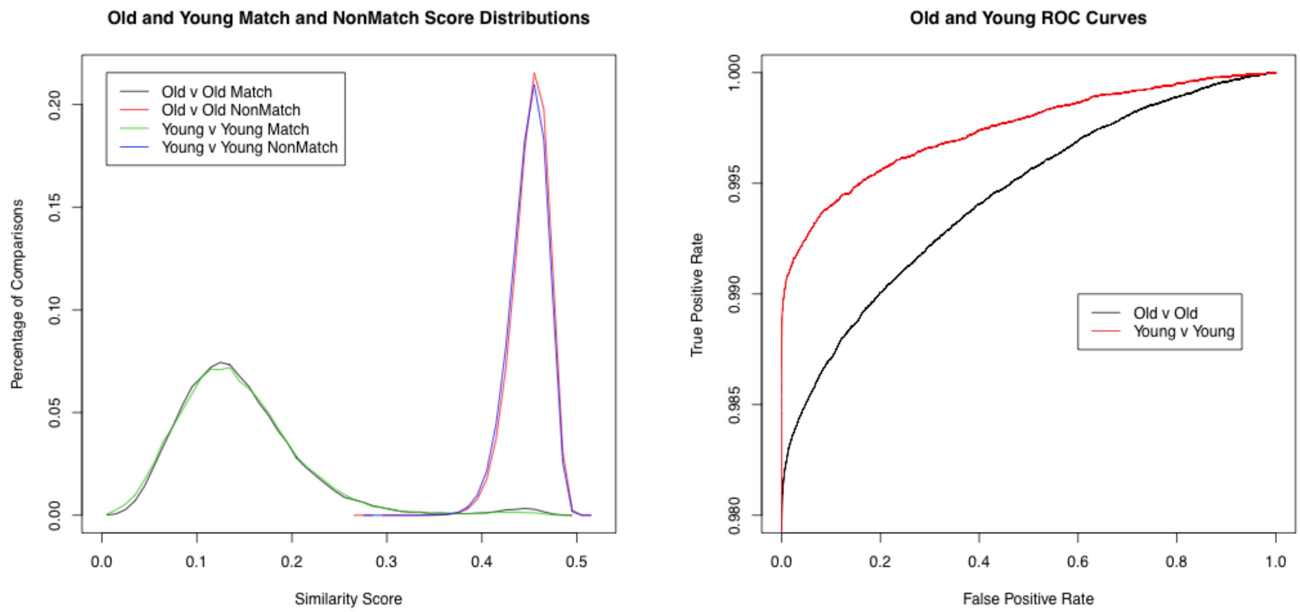


Figure 2.

Filter Number	Filter Characteristic
1	Detects small spots in texture
2	Detects large spots in texture
3	Detects thin vertical lines
4	Detects thick vertical lines
5	Detects thin horizontal lines
6	Detects thick horizontal lines
7	S5S5 Filter : Detects spot texture energy [18]
8	R5R5 Filter : Detects ripple texture energy [18]
9	E5E5 Filter : Detects edge texture energy [18]

Table 1. Description of Iris Texture Feature Filters

these regions is shown in the bottom two figures in Figure 2. Each region is then used to compute five values - average pixel intensity, standard deviation of pixel intensity, the 90th percentile pixel value, the 10th percentile pixel value, and the range of pixel intensities between the 90th and 10th



(a) (b)

Figure 3. (a) The match and nonmatch score distribution for the old and young data set all-vs-all comparison. (b) The resulting ROC Curve for the old and young comparison experiments.

percentile pixel values. Masked pixel marked in yellow in Figure 2 are not considered in these calculations. This is the same method of feature extraction used by Lagree and Bowyer in [16].

4. Experiment and Results

Subject-disjoint 10-fold cross-validation was used in this work. The data set was divided into 10 folds and each fold was evenly split with regards to the number of older and younger subjects, except for one fold having slightly fewer older images (26) than younger images (30). Balanced gender and ethnicity within folds as well as practical splits within metadata were considered.

The RandomForest algorithm as implemented by Weka using 300 trees was used to train and test a classifier based on the 630 features from each subject [14][4]. A parameter sweep was also performed across the number of trees, number of seeds, and depth of the trees to determine the best classification rate possible using this algorithm and feature set. From this experiment we found a correct classification rate of 64.68%, which we will refer to as the baseline result since all 630 features were considered.

In order to determine the statistical significance of this result a 95% confidence interval of randomness around the theoretical mean was established. The null hypothesis states that the classifier's performance is equivalent to randomly guessing the class. We first assumed our distribution to be binomial such that each trial contained 596 binary classifications, one for each image in our overall data set, which can either be true, or correctly classified, or false, incorrectly classified. The theoretical mean of this distribution would then be a correct classification rate of 0.5, or a correct classification count of 298. Our experiment yielded a correct classification rate of 64.7%, or 385.5 within this distribution. Using the normal approximation method for binomial confidence intervals, the confidence interval of randomness $[0.4, 0.5]$ or $[267.6, 328.4]$ was generated. We then see that our baseline result falls outside of this interval. Thus, we can reject the null hypothesis in favor of the conclusion that our classifier performs better than random chance. From this we may infer that there is some difference in the texture of iris images from the younger group and older group.

Two validation sets were also created in order to test this result. The first validation set was created using new images from five random subjects from the younger group and five random subjects from the older group. Each subject was represented by three new left eye images and three new right eye images, each eye pair coming from a different acquisition session. When testing our classifier using this validation set we achieved a classification rate of 71.1%.

The second validation set was generated using ten new subjects. All the subjects in this new data set are in the younger age range. No new old subjects could be added to

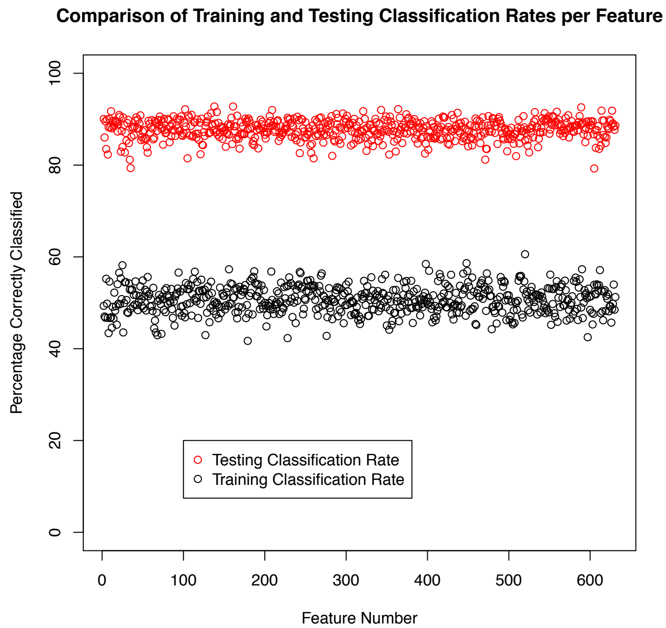


Figure 4. This graph depicts the testing and training rates of the 630 individuals features from a 10-fold cross-validation using RandomForest to classify irises from old and young subjects.

this validation set due to data pool restrictions. Each subject was represented by three left eye images and three right eye images where each eye pair was taken in a different acquisition session. When testing our classifier using this validation set we achieved a classification rate of 64.5%.

4.1. Single Feature Classification

To further investigate the performance of each of our 630 features, each feature was individually used to train and test each fold of a RandomForest classifier, using the previously discovered parameters. Figure 4 displays the results of each of these single-feature classifiers in terms of both training classification rate and testing classification rate. For a single feature, the mean training classification rate was 87.7% with a standard deviation of 2.1%. When testing each single feature classifier, the mean classification rate was 50.7% with a standard deviation of 3%. Due to the testing classification rate being approximately equivalent to random guessing, this implies that some set of features, possibly small, would improve the classification rate. This improvement may even surpass the baseline results due to the possibility that we are overfitting the data using all 630 features.

In the future we propose three methods of feature selection. The first is grounded in subject based voting results. Each subject would be tested individually using the each single classifier, producing 630 classification scores for each of the 100 subjects. Each subject would then vote for the classifier or classifiers which performed most accurately. The classifiers with the largest number of votes which also represent the most subjects would then be chosen to train a new classifier. The second method for feature selection considers a finer scale by looking at individual images, but uses the same voting procedure as the first. Here we would have 630 classification scores for each of the 596 images. The third method is the most naive, in this method each feature would be paired with every other feature. A classifier would then be trained and tested based on each pair of features. The best classification rate or rates would then decide the best pair of features to use. However, it may be the case that two features is not enough, and thus this type of experiment could be extend to triples, quadruples, or larger group of features.

5. Conclusions

In this paper we found that iris texture can be used to classify subjects by age range at an accuracy of 64%. We found this to be outside of a 95% confidence interval around random classification. It was further discovered that no individual feature performs as well as all of the features together. Thus, it is desirable to explore feature selection methods which will provide us with a set of features which best classifies that data without the possibility of overfitting, which may be seen when using all 630 features. The overall results of this experiment may imply that there are aging effects on the iris which might be used to determine, at least coarsely, the age of a given iris.

6. Future Work

This work provided a preliminary study in age prediction from iris images. By increasing the size of the data set, which would include acquiring various subjects of all ages, we could then train a classifier which could attempt to predict the age of an iris.

Further, in this work, features were chosen based on prior work in the classification of irises based on ethnicity. In future experiments, we would like to develop and explore other iris texture features which may be more relative to the effects of aging on the eye. For instance, features could be generated using pupil size, and its potential effects on the iris texture. By creating these new features we will also be able to discover the exact effects picked up on by the classifiers which best identify aging.

References

[1] S. Baker, K. W. Bowyer, and P. Flynn. Empirical evidence for correct iris match score degradation with increased time-lapse between gallery and probe matches. *International Conference on Biometrics*, pages 1170–1179, 2009.

[2] J. Birren, R. Casperson, and J. Botwinick. Age changes in pupil size. *Journal of Gerontology*, 5(3):216–221, 1950.

[3] K. W. Bowyer, S. E. Baker, A. Hentz, K. Hollingsworth, T. Peters, and P. Flynn. Factors that degrade the match distribution in iris biometrics. *Identity in the Information Society*, 2(3):327,343, 2009.

[4] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[5] C. Chen, Y. Chang, K. Ricanek, and Y. Wang. Face age estimation using model selection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop*, pages 93–99, June 2010.

[6] M. Dubbelman, V. A. D. P. Sicam, and G. L. V. der Heijde. The shape of the anterior and posterior surface of the aging human cornea. *Vision Research*, 46(6-7):993–1001, March 2006.

[7] M. Erbilek and M. Fairhurst. A methodological framework for investigating age factors on the performance of biometric systems. *Proceedings from the ACM Conference on Multimedia and Security*, pages 115–122, September 2012.

[8] M. Fairhurst and M. Erbilek. Analysis of physical ageing effects in iris biometrics. *Computer Vision*, pages 358–366, September 2011.

[9] S. P. Fenker and K. W. Bowyer. Experimental evidence of a template aging effect of iris biometrics. *WACV IEEE Workshop on Applications of Computer Vision*, pages 232–239, January 2011.

[10] Y. Fu, G. Guo, and T. S. Huang. Age synthesis and estimation via faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11):1955–1976, 2010.

[11] Y. Fu and T. S. Huang. Human age estimation with regression on discriminative aging manifold. *IEEE Transactions on Multimedia*, 10(4):578–584, June 2008.

[12] G. Guo, Y. Fu, C. Dyer, and T. S. Huang. Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Transactions on Image Processing*, 17(7):1178–1188, July 2008.

[13] G. Guo, Y. Fu, T. S. Huang, and C. Dyer. Locally adjusted robest regression for human age estimation. *IEEE Workshop on Applications of Computer Vision*, 2008.

[14] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The WEKA data mining software: an update. *SIGKDD Explor. Newsl.*, 11(1):10–18, nov 2009.

[15] K. Hayashi, H. Hayashi, and F. Hayashi. Topographic analysis of the changes in corneal shape due to aging. *Cornea*, 14(5):597, 1995.

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

[17] A. Lanitis. A survey of the effects of aging on biometric identity verification. *International Journal of Biometrics*, 2(1):34–52, January 2010.

[18] K. Laws. Rapid texture identification. *SPIE : Image Processing for Missile Guidance*, 238:376–381, 1980.

[19] X. Liu, K. W. Bowyer, and P. Flynn. Experiments with an improved iris segmentation algorithm. In *Proceedings of the Fourth IEEE Workshop on Automatic Identification Technologies*, pages 118–123, 2005.

[20] B. Winn, D. Whitaker, D. Elliott, and N. Phillips. Factors affecting light-adapted pupil size in normal human subjects. *Investigative Ophthalmology and Visual Science*, 35(3):1132–1137, 1994.