Similarity of Iris Texture Between Siblings

Will Vranderic and Kevin W. Bowyer
Department of Computer Science and Engineering
University of Notre Dame
Notre Dame, Indiana 46556 USA

wvrander @alumni.nd.edu and kwb @nd.edu

Abstract

Prior research has shown that twins' irises are similar enough in appearance for the untrained human observer to correctly determine whether an iris image pair comes from twins or from unrelated persons. We conducted a similar image pair classification study that asked participants to classify an image pair as "siblings" or "unrelated". We found that untrained human observers can classify pairs of siblings with over 57% accuracy using the appearance of the iris alone, without any proximal image content. This result is statistically greater than the accuracy of random guessing, which indicates that there is some degree of inherited texture similarity between siblings. This raises the question of how accurately one could classify siblings based on automated texture analysis.

1. Introduction

Iris recognition began in the 1990s with the work of John Daugman [1,2] and has developed into an active research area with notable successful large-scale applications [3]. Recently, studies in which humans classify iris images have shown that there is information in the appearance of the iris that is not captured by the iris codes used in iris recognition. Two such studies have been conducted, one on iris images of identical twins, and one on iris images of left and right irises of the same person. These studies have shown that untrained human observers can accurately categorize pairs of images based on perceived similarities in iris texture [4].

The implications of the twins study, in particular, prompted the experiment described in this paper. If untrained humans can identify iris images that belong to identical twins, there is a possibility that they could identify siblings as well. Should this be the case, we could imagine an automated iris texture analysis system that could detect the similarities that human observers perceive, and likewise could predict a genetic relation from a pair of iris images. The practical applications of such a system are broad. This capability would also raise new privacy concerns for current iris biometrics systems.

The rest of this paper is organized in the following manner. Section 2 summarizes related research, including a study on genetic correlations in iris texture characteristics of humans. Section 3 describes our data acquisition and image segmentation. Section 4 describes our experiment to test how much similarity humans can detect between sibling's irises. Section 5 details the results obtained. Section 6 provides a summary and conclusion. Section 7 proposes a more comprehensive study similar to this one but larger in scope, and what is needed to conduct that study.

2. Related work

Human classification studies are a new area of inquiry in iris biometrics research. Two such previous studies are particularly relevant to our experiment. The first study, conducted on twins, was able to achieve 81% accuracy for observers using only iris information, a rating that improves to 92% when the participants express confident judgment [4]. The other study, which dealt with left and right iris images of the same person, yielded similar results– 86% accuracy in distinguishing irises of the same person versus different persons, and over 93% accuracy when confident [4].

We are interested in determining if the area of study that these two experiments were concerned with can be expanded. As all of the twins in the first study were identical, both studies only tested images of genetically identical, or "monozygotic," individuals (whether from the same person or two identical twins). In this experiment, we seek to determine if humans can perceive similarities in the iris texture in pairs of images that come from (non-twin) siblings.

3. Data

The primary question that we seek to explore in this study is, "Can human observers determine whether two iris

images come from siblings at some level of accuracy better than random guessing?" We will mirror the approach taken in previous work that deals with human classification of left and right iris images of the same individuals and pairs of iris images of identical twins [4].

3.1. Acquiring data

All images used were acquired at the University of Notre Dame using an LG 4000 iris sensor [5]. We were able to obtain images for 38 sibling pairs in total: 21 male-female, 9 male-male, and 8 female-female pairs. Our experiment used both left and right eye images from these 38 pairs, and so we had a total of 76 distinct iris image pairs (152 images) representing siblings. Three of these 38 pairs came from twins, two of which were self-reported as fraternal twins, and one of which did not know. We also had two groups of three siblings, such that 6 pairs in the study share one individual with another pair. Finally, we also included 38 more pairs of unrelated people (as described in the next section), to be randomly sequenced amongst the sibling pairs in the experiments. This brings the total to 152 distinct iris image pairs (304 images).

It is important to note that sibling status is self-reported for the images acquired for this study. The self-reported siblings were not asked specific questions about whether both have the same two parents. It is possible that some self-reported siblings in our study come from "blended families" or were adopted. A future data acquisition should include more specific recording of sibling status at the time of data acquisition.

3.2. Selecting matched unrelated pairs

In order to balance the set of unrelated trials with the set of sibling trials, we attempted to select a pair of unrelated persons corresponding to each sibling pair with the same ethnicity, gender, age, and eye color. In a few cases, where we lacked the data to find an exact match, we selected a pair of unrelated individuals with different eye colors, age, or ethnicity. For example, a pair of siblings might consist of a white male with green eyes born in 1991 and a white female with brown eyes born in 1993, and the pair of unrelated individuals might be a white male with brown (instead of green) eyes born in 1991 and a white female with brown eyes born in 1993.

3.3. Pupil dilation and iris occlusion

The paper analyzing twins' iris images concluded that "it is likely easier to identify twins' irises when the irises have similar degrees of dilation," since the most misclassified twin image pairs had a noticeable difference in pupil size [1]. Because of this, we selected images such that the pupil dilation values in sibling trials and unrelated person trials were as similar as possible. Figure 1 below shows a histogram of the differences in dilation ratios across all left and right image pairs, separated into sibling and unrelated trials.

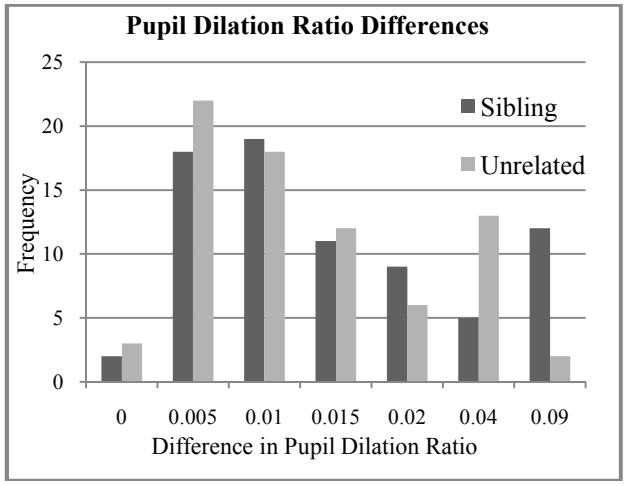

Figure 1: Histogram that shows the distribution of differences between pupil dilation ratios across all trials in the experiment, separated into sibling and unrelated trials.

3.4. Segmentation

For our experiment, we needed to accurately locate and isolate the iris. Using a helper program, we segmented the images and created new versions containing only the iris texture, leaving all other regions blacked out. In cases where the program did not segment the image correctly, we manually marked the iris and pupil boundaries of each image individually. We then re-ran the program with these new hand-marked images in order to achieve proper segmentations. By using these segmented images, with the periocular region colored black, we hoped to provide the most accurate answer to our primary question, of whether or not participants can correctly identify siblings from only the appearance of the iris texture.

4. Experimental procedure

Before the actual experiment began, the participant was given a few help pages that explained the experiment. Following these help pages, ten trial image pairs (five of siblings of which four were from identical twins, and five of unrelated persons) were then displayed so that the participant could get a feel for how the experiment would proceed. This also gave them a chance to see what a sibling pair looked like and what an unrelated pair looked like. The timing and format of the test image pairs was the same as that of the actual experiment pairs, and is described next.

After the participant continued past the training image pairs, the remaining 152 image pairs, which constitute the

actual experiment, were presented. Each participant had a different randomly generated order for viewing the 152 image pairs. Each image pair was displayed side by side for an unlimited amount of time. Underneath the image pair, the participant selected one of five options to indicate whether they think the pair comes from siblings, or from unrelated persons:

1. Certain that these images ARE of two siblings
2. Likely that these images ARE of two siblings
3. Don't know
4. Likely that these images are NOT of siblings
5. Certain that these images are NOT of siblings

After the user selected an option, the display changed to tell the user whether their selection was correct or incorrect. After the participant pressed next from this feedback slide, the next pair of images was displayed. The study continues in this way until all pairs of iris images were displayed. The final slide displays the number of image pairs correctly identified (between 0 and 152) over the total number of image pairs in the experiment (152).

5. Experimental Results

Twenty individuals in total participated in our experiment. However, one of these individuals selected "Don't Know" 29 times, representing approximately 20% of the total trials. The participant with the next highest number of trials marked as "Don't Know" totaled less than 5% of the 152 trials (7/152). Thus we decided that results from the individual with 29 trials marked as "Don't Know" effectively represent an outlier relative to the other participants' results, and should not be included. Unless otherwise noted, numbers and averages in the following results subsections do not include this individual's data. After excluding the outlier participant, the average across all other participants was just over 1% of the total trials (1.68/152).

As mentioned previously, we allowed each image pair in the experiment to be displayed for an unlimited amount of time. The two related studies mentioned earlier only allowed participants to see each image pair for a limited amount of time before they had to make a judgment on whether the image pair was from twins or identical persons. Figure 2 below shows the total time for each participant to complete our experiment and the corresponding accuracy. There does not appear to be a noticeable correlation between time each participant needed for the experiment and the accuracy that the participant achieved.

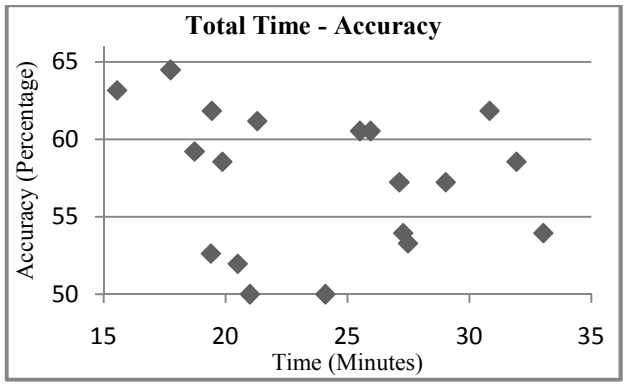

Figure 2: Scatter plot that shows the time each of the 19 participants needed to complete the experiment against the accuracy that each participant achieved.

5.1 Can humans tell siblings from iris texture alone?

To calculate an overall accuracy score, we counted the total number of times each participant answered "likely" or "certain" of the correct response of siblings or unrelated. In other words, we did not distinguish based on the tester's confidence level, but only whether they believed the image pair was from siblings when they were in fact from siblings, or whether they believed the pair was unrelated when they were in fact unrelated. We then divided the total number of correct responses by 152, the total number of image pairs in the experiment. We then averaged all participants together to obtain a mean score of 57.4% for the experiment.

The minimum score for a particular subject was 50%, and the maximum score was 64.5%. The maximum score that any person scored on the subset of unrelated person questions was 71%, and the minimum was 40.8%. For the subset of sibling pair questions, the maximum score that a subject achieved was 80.3%, and the minimum score was 43.4%.

We took two different approaches to determine if our results were statistically significant. First, we used a standard t-test to evaluate the null hypothesis that our human participants did not perform differently than random guessing when identifying iris image pairs as siblings or unrelated. The resulting p-value was less than 10^{-4}, hence we have statistically significant evidence that our testers were distinguishing siblings and non-siblings more accurately than if they were randomly guessing. Our second approach was to calculate a binomial probability distribution for each individual participant's results. From this we found that 12 of the 19 participants scored significantly better than 50% on the total 152 trials. We can conclude from both approaches that

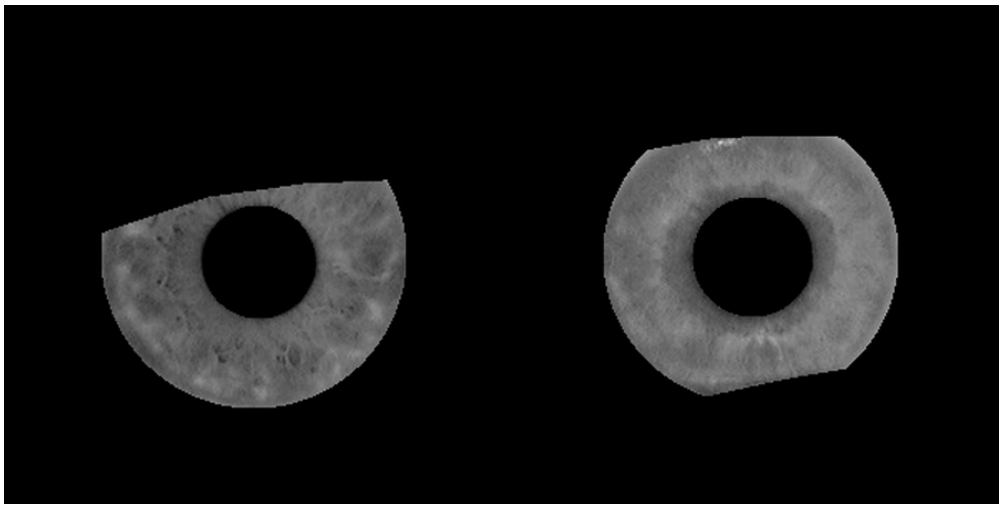

Figure 3: Right irises of the male-female pair that was identified correctly as unrelated by all 19 participants.

participants are able to detect similarities between iris images from two siblings at accuracy greater than random guessing.

5.2 Which image pairs were most frequently classified correctly and incorrectly?

Eight trials were categorized correctly by all subjects. Three of these eight were sibling trials and five were unrelated person trials. The three sibling trials represented two brother-brother trials and one brother-sister trial. Four of the unrelated person trials that were categorized correctly by all subjects represented male-female image pairs, and the other was a male-male image pair. Two of these four trials were the left and the right irises of the same male-female pair. This mixed-gender unrelated matched pair was identified correctly in both questions, the left and right irises, by all participants. The right irises of this unrelated pair are shown above in Figure 3.

Figure 4 shows the right iris images of the most correctly identified sibling pair, a male-male pair. All participants identified the right iris images correctly, and all but two participants identified the left iris images correctly (not shown).

One trial was categorized incorrectly by all subjects. This was a brother-brother sibling pair trial and is shown in Figure 5. The overall worst scoring sibling pair had only two participants correctly identify the left images and three correctly identify the right images. Similarly, the worst unrelated pair had only two people guess the right and left images of this mixed gender pair.

5.3 Are sibling or unrelated pairs more difficult?

When considering the scores over all iris images, it seems

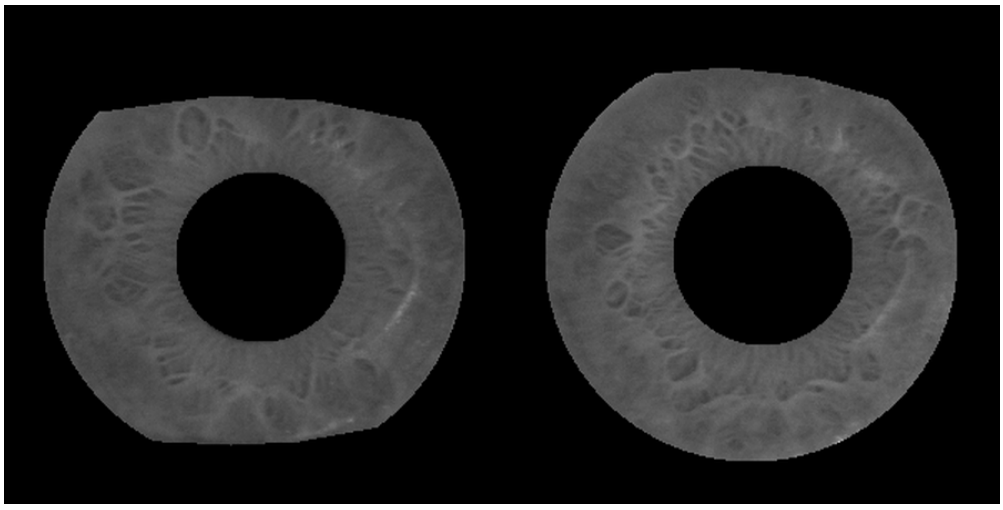

Figure 4: Right irises of the male-male pair that was identified correctly by all 19 participants as siblings.

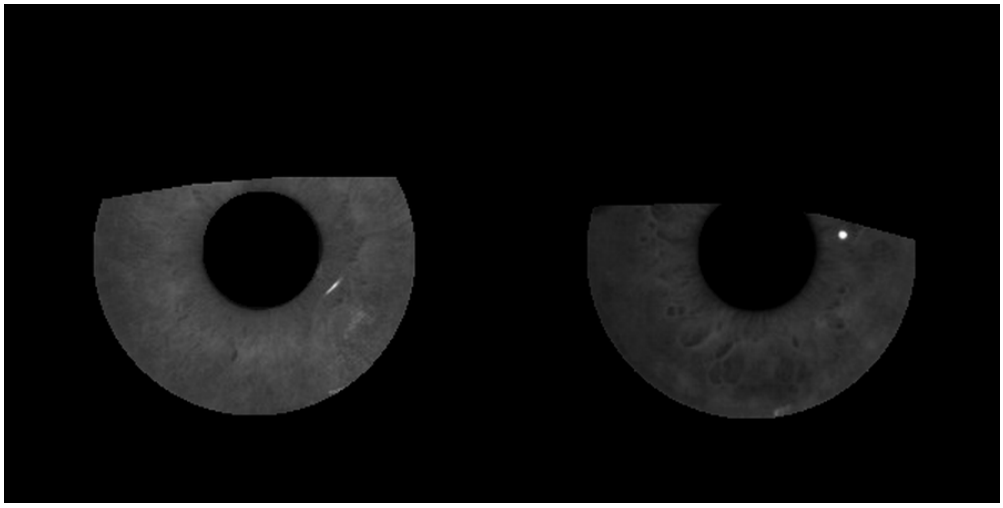

Figure 5: The male-male sibling pair whose right iris images were identified incorrectly by all 19 participants.

to be about equally difficult to label siblings as siblings and to label unrelated people as unrelated. There were 76 total question pairs of siblings and 76 of unrelated persons. The average percentage correct on the 76 unrelated image pairs was 57.2%, and the average score on the 76 siblings' pairs was 57.6%.

5.4 Did subjects learn during the experiment?

There was a slight improvement in the average number of images pairs categorized correctly from the first half of the experiment to the second half. The average on the first half of the experiment was a 56.2%, whereas the average on the second half was a 58.5%. Analyzing the results in terms of first, middle and last thirds yields a similar result. The fraction of image pairs categorized correctly in the first third of the experiment was 56.5%, the fraction for the middle third was 57.8%, and the fraction for the final third was 57.9%. The improvement seen in the later portion of the experiment in either case is only slight, but it may suggest that experience and training can play a role in a larger human classification study similar to this one.

5.5 Were subjects confident in their classifications?

Considering the trials as sibling pairs and unrelated image pairs, 1229 (85.6%) of the 1435 siblings responses were classified with the "likely" option, and the other 206 (14.4%) responses were ones where participants said that they were "certain" that the image pair was from unrelated persons. The breakdown of the answers where participants thought the pair was from unrelated persons was similar - 1181 (80.1%) of the unrelated person responses were made with the "likely" option, with the remaining 240 (16.3%) of these selections being made with the "certain" option. The fact that the participants were much more likely to judge their decision as "likely" than as "certain" reinforces the point that the experimental task was difficult. This is certainly the case in comparison to the studies reported in [4]. A bar chart of the frequency of the different responses appears in Figure 6. A bar chart of accuracy by response type appears in Figure 7.

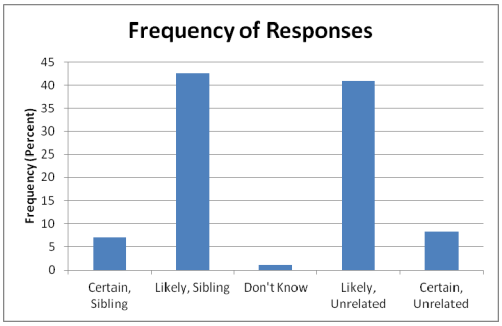

Figure 6: Overall frequency of responses in the experiment.

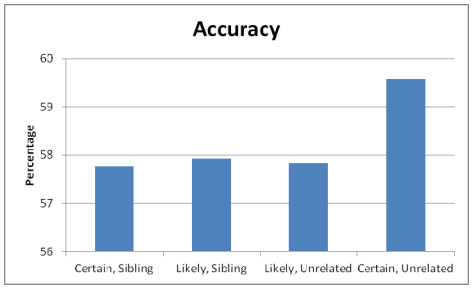

Figure 7: Accuracy by type of response.

6. Conclusion

We have found that siblings have a level of iris texture similarity that allows better than random classification of image pairs. When presented with unlabeled sibling and unrelated image pairs, humans can classify pairs of siblings with 57.4% accuracy using only the appearance of the iris.

The problem studied in this experiment seems to be quite hard. Previous studies on classifying monozygotic irises used limited time per trial, and subjects performing classification had no prior experience at the task, and the overall accuracy was still over 80% correct. In this experiment, subjects had unlimited time to consider each trial, and overall accuracy was only about 57%.

We looked at whether or not participants performed better when they were confident (marked "certain" instead of "likely") in their selection. As mentioned earlier, the related twins study achieved 81% accuracy, which improved to 92% accuracy when the participants expressed confident judgment [4]. In the study in this paper, however, there was only a slight improvement from 57.9% accuracy when participants were not confident, to 58.7% when they were confident. We believe this stems partially from the fact that participants selected with certainty in only 15.4% of responses, as compared to the much higher percentage of confident responses in the two related studies of around 30% and 45% [4].

We found that same-gender sibling pairs might be easier to classify than mixed-gender pairs. Participants had a 60.1% accuracy rating for sibling pairs that were of the same sex, as opposed to a 55.5% accuracy rating for sibling pairs of mixed gender. When identifying siblings that were male-male pairs, participants had a 59.1% accuracy rating; for female-female sibling pairs, participants achieved 61.2% accuracy.

We also looked at whether or not human observers may be able to learn how to classify iris image pairs as siblings or unrelated. The accuracy improvement seen in the later portion of the experiment is slight, but it may be interesting to see if training can play a role in a larger human classification study similar to this one.

We recognize that the scope of this experiment was limited, as only twenty individuals in total participated in rating the image pairs and only 38 different pairs of siblings' irises were used. We intended for this study to provide a basis to gauge whether further data collection of siblings was warranted for a larger, more complete experiment. We conclude that a comprehensive study of this kind, which is detailed in the following section, is needed.

7. Future Work

From the results we have obtained from this study, it appears that a larger study is worthwhile. In particular we want to further explore the apparent similarities of same-gender sibling pairs. The main thing that we need to carry out this study is a larger pool of sibling iris images. While the mixed-gender sibling pairs obtained some positive results, the same-sex pairs seemed to be easier to classify as siblings. However, we had very limited numbers of these same-sex pairs; we had more than twice as many mixed-gender pairs (21) as either same-gender pair category (9 male-male and 8 female-female). We need more sibling pairs in general to conduct a larger study, and we particularly need more same-gender sibling pairs.

Aside from a larger version of this particular experiment, we also might consider looking into other familial relations. A parent-child study might be beneficial both as a new study and possibly to confirm that the apparent similarity in sibling iris texture has to do with the familial relation of the pair.

Finally, we think it would be interesting to see if we could train a classifier that would accurately be able to determine if a pair of images comes from siblings or from unrelated persons. We would compute feature vectors from the image pairs and see if a classifier could use these to build and improve a decision tree to accurately determine if the pair comes from siblings.

References

[1] J. G. Daugman, Biometric personal identification based on iris analysis, U.S. Patent 5,291,560, 1994.

[2] J. Daugman, Recognizing persons by their iris patterns, Chapter 5 in Biometrics: Personal Identification in Networked Society, A. K. Jain, R. Bolle, and S. Pankanti, eds, Kluwer, Norwell, MA, 1998.

[3] M. J. Burge and K. W. Bowyer, Handbook of Iris Recognition, Springer, 2013.

[4] K. Hollingsworth, K. W. Bowyer, S. Lagree, S. P. Fenker and P. J .Flynn, Genetically Identical Irises Have Texture Similarity That Is Not Detected By Iris Biometrics, Computer Vision and Image Understanding 115 (2011), 1493-1502.

[5] www.lgiris.com/ps/products/irisaccess4000.htm, accessed July 5, 2013.