A Longitudinal Study of Iris Recognition in Children

Morgan Johnson
Clarkson University
Potsdam, NY USA
johnsomi@clarkson.edu

David Yambay
Clarkson University
Potsdam, NY USA
yambayda@clarkson.edu

Dan Rissacher
Clarkson University
Potsdam, NY USA
drissach@clarkson.edu

Laura Holsopple
Clarkson University
Potsdam, NY USA
lholsopp@clarkson.edu

Stephanie Schuckers
Clarkson University
Potsdam, NY USA
sschucke@clarkson.edu

Abstract

Biometric recognition is dependent on the permanence of the biometric characteristics over long periods of time. However, there has been limited research in this area, particularly in children during development. This paper presents the start of a longitudinal study of irises in children to understand when biometrics can be used reliably and the effect aging has on the biometric modality as a child grows. Data was collected and analyzed in children ages 4-12 years over three visits, spaced approximately six months apart. This is one of the few iris collections spanning this broad age range in children. The results show that there is a slight decrease in match scores between the resultant comparison of collection 1 to collection 3 (12 months difference) and the resultant comparison of collection 1 to collection 2 (6 months difference); analysis shows this difference is not statistically significant. Additionally, the data analyzed resulted in very similar iris recognition performance when examining a subset of subjects in fifth grade and a subset of subjects in first grade. These results could indicate that the iris biometric characteristic is stable over time, at least as early as age 4, the youngest group tested in this work. Additional longitudinal data is needed to support this hypothesis.

1. Introduction

The study of biometric recognition has gained interest to support various applications such as immigration, refugees, and distribution of benefits. There is a desire to also utilize biometric recognition for children to support these applications. However, it is unclear whether biometric characteristics change as a child ages. Research topics related to this include the determination of the earliest age that a biometric modality is viable, how a modality varies with aging, and the development of models to account for variations of a modality over time. Furthermore, if there is an effect of aging on biometric modalities, there could also be potential indicators of age inherent to modalities, whereby a biometric modality could be used to estimate age, creating a tool for a variety of applications. This paper focuses on iris recognition in children between the ages of four to twelve years. The goal of this research is to analyze the variability of iris recognition in children over time. This paper focuses on the three collections taken thus far and analyzes iris recognition performance, comparing iris images from the first collection to subsequent collections.

2018 IEEE 4^{th} International Conference on Identity, Security, and Behavior Analysis (ISBA) 978-1-5386-2248-3/18/$31.00 *©2018 IEEE*

2. Background

Currently, the field of aging biometrics in children has not been heavily researched. One of the first studies showed how the effects of aging when regarding facial appearance can be explained when utilizing a parameterized statistical model [1]. Many geometric transformations have been proposed to model craniofacial growth, including rigid rotation, affine shear, and cardioidal strain [e.g., 6].

Other studies focusing on fingerprint in children have looked at the stability of the line pattern's structure of a fingerprint [2]. Intensive pediatric research established that a fingerprint's pattern is developed and finalized at an estimated age of 24 weeks [3]. A significant longitudinal study assessed fingerprint recognition in 309 children from ages 0 to 5 over a one year period [4, 5].

It is known that a person can be identified through the pattern in their iris, but the permanence of this measurement is under study [9], [10]; however, this research focuses on adults. One study looked at the viability of using commercial sensors to capture images of children's irises [7]. They found that the recognition accuracy for the enrolled subjects was very high, but did not investigate further due to high failures in enrollment. To the best of our knowledge there are no longitudinal studies of growth-related effects on children's irises.

3. Experiment

Researchers worked with a local elementary school to identify participants for this study, while following an approved IRB protocol. The researchers were granted permission to utilize a room in the elementary school where collections could be held. There have been three collection events thus far; each collection is separated by approximately 6 months. Six modalities were collected, including the subject's iris. This paper will focus on the iris data taken from these collections and will analyze the data to determine if the pattern changes over time.

Table 1 shows the number of participants during each collection. There were 171 participants in collection 1, 179 participants in collection 2 and 172 participants in collection 3.

Table 1. Number of subjects that participated per collection.

Collection Number	Participants
1	171
2	179
3	172

Table 2 shows how many subjects have data from collection 1 and collection 2 as well as how many subjects have data from collection 1, collection 2, and collection 3. A total of 123 subjects have data from collection 1 and collection 2, while only 114 subjects have data from all three collections. This paper will only analyze the 114 subjects that were able to successfully complete all three collections.

Table 2. Number of subjects that have data from multiple collections.

Collection Number	Participants
1, 2	123
1, 2, 3	114

The difference in participants include a child being absent from school the week the collections are taking place at the elementary school or that the child's family has moved out of the school district. Additionally, every new school year the study will add on a group of children, typically, around 4 years old. Finally, it should be noted that iris image collection in the very young ages (4 or 5) was sometimes not possible due to attention issues, an inability to stay still, or an unwillingness to have their data taken. While we were not able to collect iris images in some children, this is an issue of usability and is not the subject of this study.

3.1. Subject Enrollment

This research was conducted under an approved IRB protocol. Parents are required to sign an informed consent document and the child is required to give an age-tiered assent ranging from verbal to signed consent. A unique subject number is affixed to the consent form, and is used throughout the study to identify the individual.

3.2. Dataset

The iris portion of the child biometric dataset used in this paper was collected using an IG-AD100 Dual Iris Camera manufactured by IrisGuard, a commercial iris sensor software for capture of the irises. The flashing lights automatically control pupil dilation, reducing noise in the collection. The camera also utilizes auto-focus, but this is limited to a window of distances. As such, the child is instructed to remain still and the collector adjusted the camera distance as needed. Additionally, a chin rest, similar to ones used in eye examinations, was used to help keep the child stable and aligned with the camera as shown in Figure 1. Additionally, the blinds are closed in the collection room to create a constant environment.

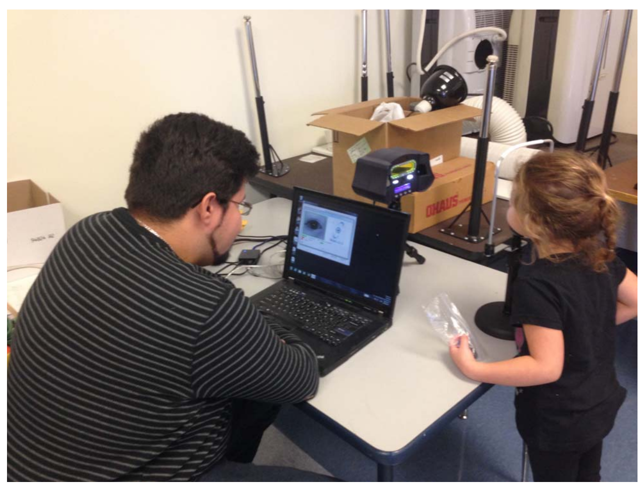

Figure 1. Demonstration of Iris collection using IrisGuard system and chin rest for stability/alignment in children.

At the beginning of the collection process (collection 1) only two images were taken of each eye. For the subsequent collection, it was decided that four images of each eye should be taken in order to increase number of images and support the data analysis (collections 2 and 3). This means that there are four images of each eye taken per collection, resulting in twenty four images taken in an ideal situation.

3.3. Example

Below is an example of a two subjects' biometric data from all three collections. The first collection was taken on January 4, 2016. The second collection was taken on October 26, 2016. The third collection was taken on April 25, 2017. These collections are approximately six months apart. Figures 2 and 3 are examples of the data that are used for data analysis.

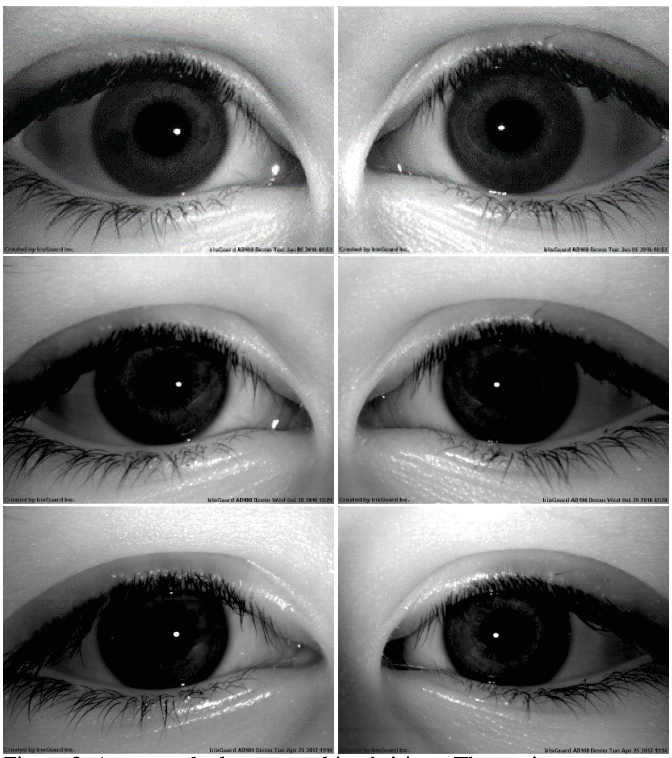

Figure 2. An example the same subject's irises. The top images are from collection 1 (0 mos), the middle from collection 2 (6 mos), and the bottom from collection 3 (12 mos).

Only one image from each collection is shown in Figures 2 and 3. The first image taken in collection 1 is used as the enrollment image for all comparisons and is consistent throughout all of the data analysis.

3.4. Verieye Matching Software

To analyze the change in a child's iris over time the commercial product Verieye was used. Templates were created using Verieye software for each subject for collection 1, collection 2, and collection 3. All analysis was performed on left eye images. Right eye images have not been analyzed to date. The selected image is compared to the entire database and a score is outputted for each. The maximum score is 1557. This score will appear if an image is compared to it-

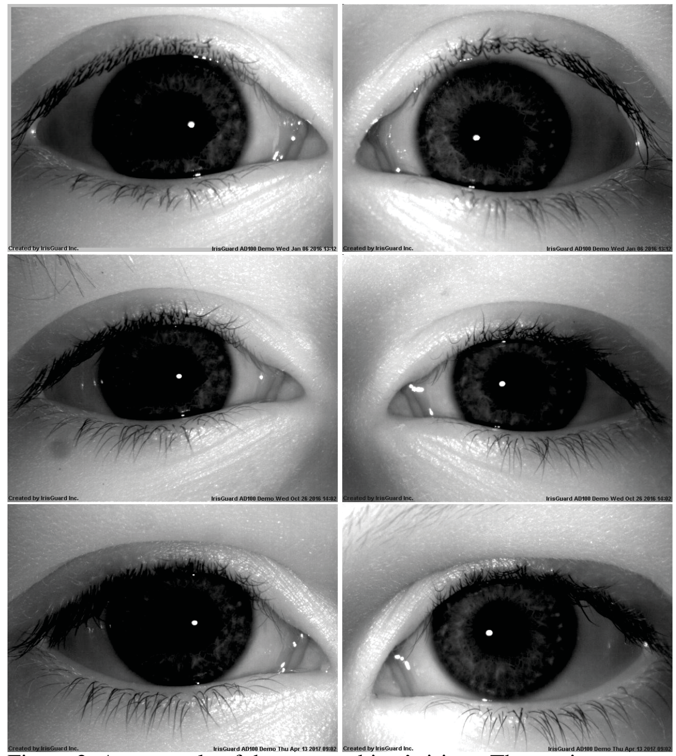

Figure 3. An example of the same subject's irises. The top images are from collection 1 (0 mos), the middle from collection 2 (6 mos), and the bottom from collection 3 (12 mos).

self. Verieye outputs a score of 0 when it rejects two irises as a match.

4. Data Analysis

In the first round of analysis the first image of the left eye in collection 1 was compared to the entire database of templates. In an ideal situation, this template would match with the subject's other left eye images from collection 1, the subject's left eye images from collection 2, and the subject's left eye images from collection 3. Collection 1 compared to collection 1 should produce high scores since the images were taken within milliseconds of each other without a change in environment. When comparing collection 1 to a collection on a different day a lower match score is expected, as it is for most biometric modalities. Although the environment stayed constant – same room, closed blinds, same machine and process – it is expected that there will be more variability between days which include such factors as pupil dilation, lighting, and chin position.

This subject's probe is also compared to the entire database of all subjects to check for false accepts. It should be noted that not all subjects were able to complete all three collections. As mentioned, for collection 1 only two photos were taken of the child's left eye, while four photos were taken for collections 2 and 3. This means for each probe, Verieye should ideally come back with ten scores, two from collection 1, four from collection 2, and four from collection 3 for left eyes only. Comparisons with itself are discarded. All comparisons are with collection 1. This paper will refer to the comparisons as 1-1, 1-2, and 1-3 where 1-2 has a six month gap between images and 1-3 has a twelve month gap.

For 1-1 comparisons, the results were averaged per subject, excluding the 1557 score. The 1557 score was excluded because this score only represents the probe matching to its own template, making it a check point that the template is correct, but not a meaningful data point. For 1-2 comparisons, all four scores were averaged, and all four scores of 1-3 were averaged.

The first collection was completed in the spring of 2016, collection 2 in the late fall of 2016, and collection 3 in April of 2017. The graph below in Figure 4 shows only five subjects out of 114 with complete data from all three collections. This graph is for visual reference.

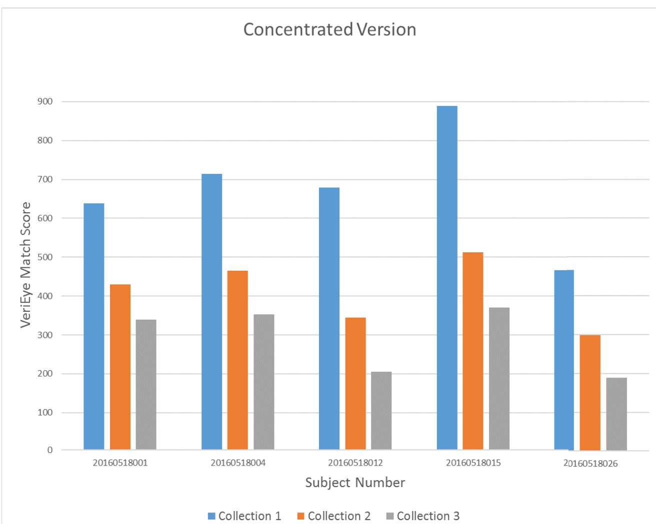

Figure 4. An example of five randomly selected subjects and the average Verieye match scores for each collection

Figure 5 shows the averages of the 1-1 comparisons, 1-2 comparisons, and 1-3 comparisons by subject number. It should be noted that in the beginning of the collection process, it was difficult to capture more than one image of an iris. Thus, some subjects only have a single left eye image from collection 1. This image was used for 1-2 comparisons and 1-3 comparisons, but do not have a 1-1 comparison score. This results in fewer points than 114 for the 1-1 comparison graph.

Table 3 and Figure 6 provide the average of the 1-1 comparisons, 1-2 comparisons, and 1-3 comparisons, as well as the standard deviations. The Table also shows the maximum and minimum scores for each comparison. The average of 1-1 comparisons is 790.3, the average of 1-2 comparisons

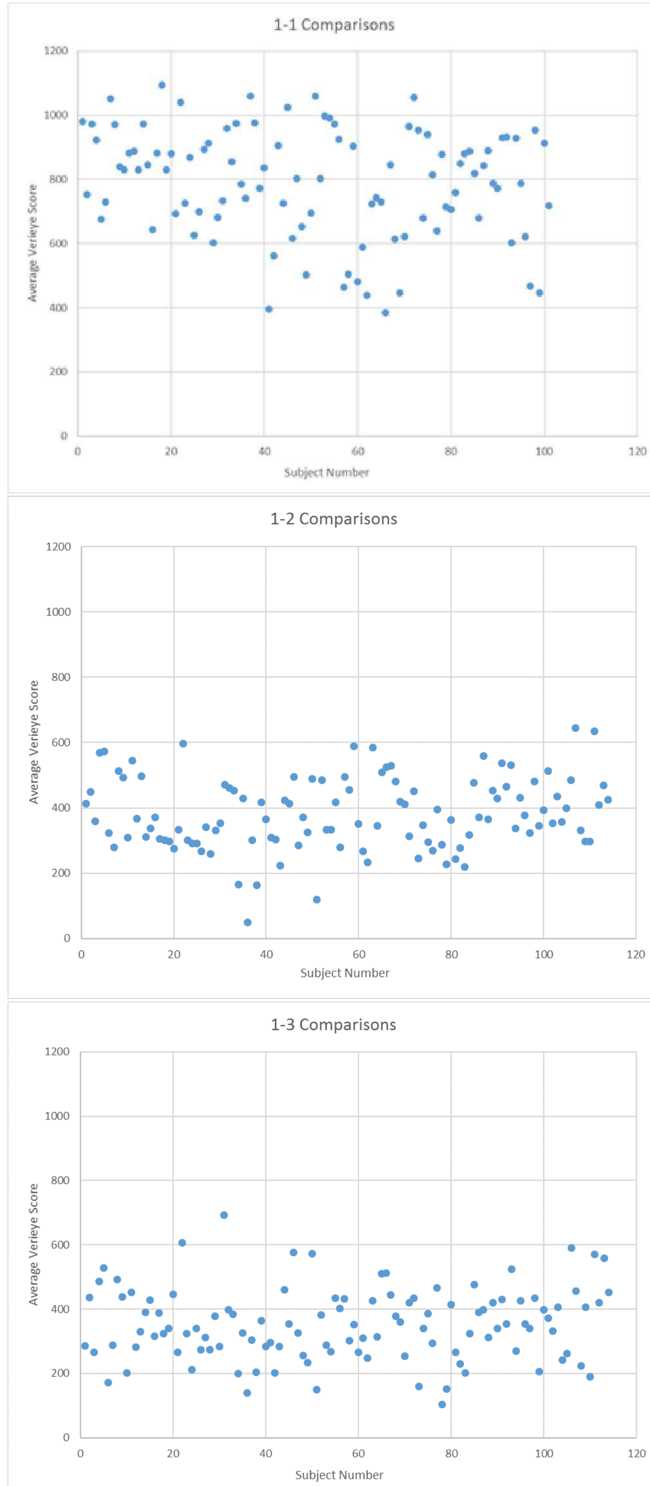

Figure 5. Scatter plots showing all subjects with complete data sets and their average Verieye Score comparisons for each collection. The first graph is the averages of the 1-1 comparisons, the second graph is the averages of the 1-2 comparisons, and the third graph is the averages of the 1-3 comparisons.

is 380.6, and the average of 1-3 comparisons is 352.5. The standard deviation of the 1-1 comparisons is 169.1, while the standard deviation for the 1-2 comparison and the 1-3 comparison are approximately 111.

A statistical analysis of the difference between the two averages of 29.1 has a t-difference of 1.976. This results in a 95% confidence interval that there is no significant difference between the means of the two populations.

Table 3. The average, standard deviation, maximum, minimum and percent difference for each comparison.

Comparison	**1-1**	**1-2**	**1-3**
Average	790.3	380.6	352.5
Standard Deviation	169.1	110.5	111.9
Maximum Score	1377	652	698
Minimum Score	310	50	124

As shown, the 1-1 comparison scores are much higher than the 1-2 and 1-3 comparisons, as expected, due to highly correlated nature of a single visit. Additionally, the 1-2 comparison scores are higher than 1-3 comparison scores. To compare the percent difference, the 1-3 average was subtracted from the 1-2 average and then divided by the 1-2 average and resulted in a percent difference of 7.38%.

It should be noted that Verieye had one false accept of a score of 49 with a subject in collection 3 and zero false rejects for this group of subjects.

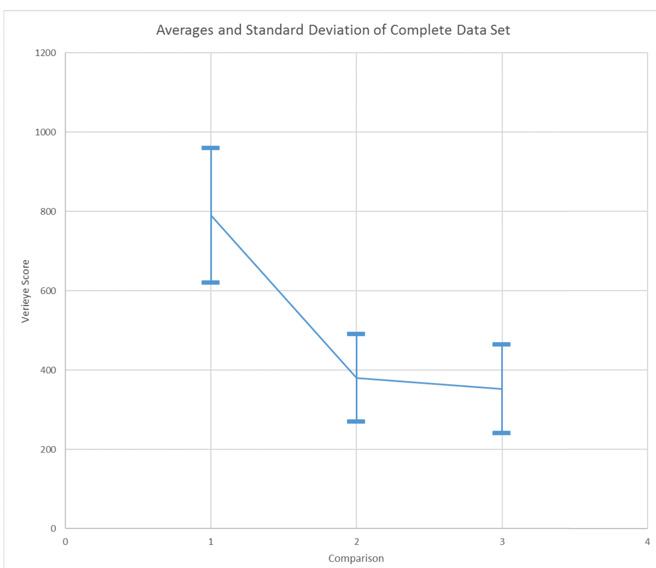

Figure 6. The averages of each comparison as well as one standard deviation away from that average for images from first collection compared to second and third collections.

4.1. Age Analysis

The next analysis compares subsets of data based on age. For this analysis two groups of subjects were used. In the original consent process the subject's birthday was asked on a volunteer basis. Some consents forms did not provide this information. For this reason, the data analysis will utilize the child's grade. The first group are children that were in fifth grade during collection 3 and the second group are children that were in first grade during collection 3. Fifth graders will be around the age of 10 and 11, while first graders will be around the age of 6 and 7. Figure 7 is a graph that shows the subset of 5th graders averages per comparison. There are 14 subjects in this subset.

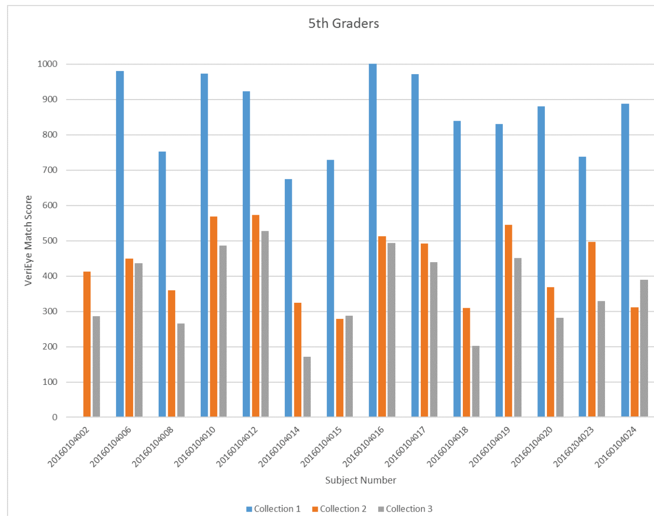

Figure 7. The group of 5th graders with data from all three collections and their average Verieye Score for each comparison.

Table 4 shows each comparison's average and standard deviation. Again the 1-1 comparison was significantly higher than the 1-2 and 1-3 comparisons, as expected. The 1-2 comparison was slightly higher than the 1-3 comparison. The 1-2 comparisons had an average of 428.5, and the 1-3 comparisons had an average of 360.4. All three comparisons had a standard deviation around 100. Finally, the percent difference between the 1-2 and 1-3 comparison was 15.9%.

Table 4. The average and standard deviation for each comparison for the subset of data including only subjects in fifth grade.

Comparison	**1-1**	**1-2**	**1-3**
Average	864	428.6	360.4
Standard Deviation	115.6	104.0	114.3

The second subset was the subjects that were in first grade during collection 3. Figure 8 shows these subjects and their average score per comparison. This subset includes 12 subjects. This graph looks very similar to all other graphs made thus far, following the trend of a high 1-1 comparison and the 1-3 comparison being the lowest average of the three.

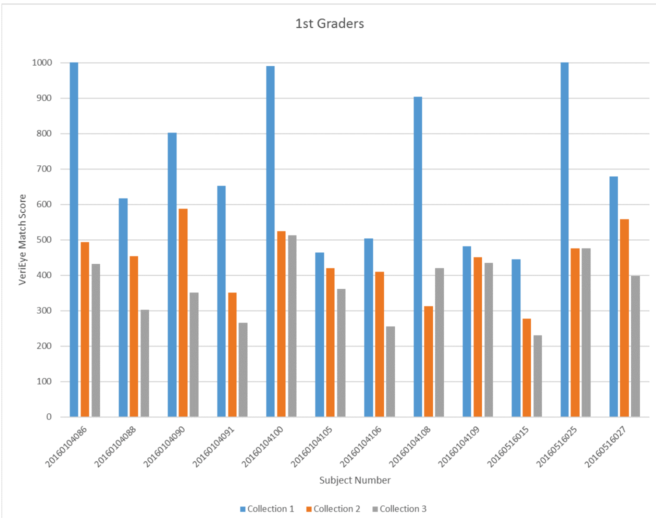

Figure 8. Subjects in 1st grade with data from all three collections and their average Verieye Score for each comparison.

Table 5 shows each comparison's average and standard deviation for the subset of 1st graders. Again the 1-1 comparison was significantly higher than the 1-2 and 1-3 comparisons and the 1-2 comparison was slightly higher than the 1-3 comparison. The 1-1 comparisons had an average of 718.4, the 1-2 comparisons had an average of 443.3, and the 1-3 comparisons had an average of 369.9. The 1-1 comparisons had a high standard deviation of 229.9, while the 1-2 and 1-3 comparisons had standard deviations around 93. Finally, the percent difference of the 1-3 comparison and the 1-2 comparison is 16.5%.

Table 5. The average and standard deviation for each comparison for the subset of data including only subjects in fifth grade.

Comparison	**1-1**	**1-2**	**1-3**
Average	718.4	443.3	370.0
Standard Deviation	229.9	95.0	91.3

When comparing Table 4 to Table 5 it can be seen that these tables have very similar values. The averages, standard deviations, and percent difference between the 1-2 and 1-3 comparisons are all similar for the subset of subjects in 5th grade and the subjects in 1st grade.

4.2. Conclusion

The results in children, grades pre-K to 5, show a slight decrease in match scores, which is not statistically significant, when comparing two irises which were collected with a time difference of 6 months and 12 months. It should be noted that the environment could affect the iris comparisons, such as a difference in the lighting. The study was performed in the same room with the blinds closed in order to minimize the impact of environment. However, it is difficult, if not, impossible to eliminate other factors which may decrease performance.

Additionally, no false rejects were seen, showing no operational impact between the comparisons of 6 month and 12 month time differences. This could indicate that this modality may be viable as early as 4 years old. While reasonable iris recognition performance has been observed to date, with the span of 12 months and pool of only 114 subjects, we are unable to make predictions of long-term large-scale performance at this stage of the research.

Finally, a comparison of the subsets between the subjects in first grade and the subjects in fifth grade show similar trends and variability.

In conclusion, the analysis of this data set did not result in a statistically significant decrease between the match score comparisons of two iris images with a six-month time window compared to a twelve-month time window. This study supports the conclusion that a child's iris does not change significantly in the span of one year. Additional research should evaluate this with a larger subject pool and an increase in the length of time of the study.

References

[1] Lanitis, A, Taylor, CJ, and Cootes, TF. "Modeling the process of ageing in face images." Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on. Vol. 1. IEEE, 1999.

[2] Gottschlich, C, Hotz, T, Lorenz, R, Bernhardt, S, Hantschel, M and Munk, A, "Modeling the growth of fingerprints improves matching for adolescents." Information Forensics and Security, IEEE Transactions on 6.3 (2011): 1165-1169.

[3] "Quantifying the Dermatoglyphic Growth Patterns in Children through Adolescence." Report NJI contract 2005-DD-BX-K056, UltraScan Corp, 2010.

[4] Jain, AK, Cao K, and Arora SS. "Recognizing infants and toddlers using fingerprints: Increasing the vaccination coverage." Biometrics (IJCB), 2014 IEEE International Joint Conference on. IEEE, 2014.

[5] Jain, AK, Arora, SS, Cao, K, Best-Rowden, L, & Bhatnagar, A. Fingerprint Recognition of Young Children. IEEE Transactions on Information Forensics and Security, 12(7), 1501-1514, 2017.

[6] Ramanathan, N, Chellappa R, and Biswas, S. "Computational methods for modeling facial aging: A survey." Journal of Visual Languages & Computing 20.3: 131-144, 2009.

[7] Schumacher G, "Fingerprint Recognition for Chil-

dren." Joint Research Centre Tech Report, Fingerprint recognition for Children, 2013.

[8] Ramanathan, N, and Chellappa, R. "Modeling age progression in young faces." Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. Vol. 1. IEEE, 2006.

[9] Baker SE, Bowyer KW, Flynn PJ, Phillips PJ. "Template Aging in Iris Biometrics. In: Burge M., Bowyer K. (eds) Handbook of Iris Recognition." Advances in Computer Vision and Pattern Recognition, 2013.

[10] Grother PJ, Matey JR, Tabassi E, Quinn GW, and Chumakov M. "IREX VI-Temporal stability of iris recognition accuracy" NIST Interagency/Internal Report (NISTIR)-7948 (2013).