2018

A Longitudinal Analysis on the Feasibility of Iris Recognition Performance for Infants 0-2 Years Old

Torrey Hutchison
thutchi@purdue.edu

LONGITUDINAL ANALYSIS ON THE FEASIBILITY OF IRIS RECOGNITION PERFORMANCE FOR INFANTS 0-2 YEARS OLD

by

Torrey Hutchison

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science

Department of Technology Leadership & Innovation

West Lafayette, Indiana

August 2018

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Stephen J. Elliott, Chair

Department of Computer and Information Technology

Kevin J. O'Connor

Department of Computer and Information Technology

Dr. Kathryn C. Seigfried-Spellar

Department of Computer and Information Technology

Approved by:

Dr. Patrick Connolly

Head of the Graduate Program

I dedicate this this in loving memory of Ruth Miller, my great-grandmother who passed shortly after I began graduate school. You have touched so many lives, from teaching high school students, night classes to adults, inmates who went on to get their GEDs, not to mention the unsurmountable help you offered, to anyone and everyone, even when you didn't have much to give. I will always remember the influence you had on my life and how your values, that have been passed down through generations, have shaped me into the person that I am today.

ACKNOWLEDGMENTS

As much as I wanted it to, this thesis did not write itself. I cannot even begin to describe the amount of educational, psychological, and academic support that I received during grad school and how much that support fueled my writing, my motivation, and eagerness to pursue a unique, beneficial, and rewarding research topic.

First, I would like to thank my committee chair Dr. Stephen Elliott - Throughout my academic career you have been not only an advisor but a mentor and a friend. I am thankful for the opportunities that you have given me; the encouragement and advice to pursue a topic that I am passionate about; and most of all, your continuous feedback, which not only propelled and strengthened my research, but fueled my appetite to continuously learn and improve. Without your passion and ability to bring out the best work in all of those around you, I would not have pursued further education, and I most certainly would not of had the courage to grit my teeth, push forward, and see this research through.

Kevin O'Connor – Without your support, even in the most trying times, this research would not be possible. I would like to thank you for all the advice you have given me and all the effort and time you put into collecting data. I will always remember the challenges and experiences we had during our infant data collection. Dr. Kathryn Seigfried-Spellar, you have been an invaluable addition to my committee. Your knowledge, recommendations, and attention to detail has strengthened this research and made this work possible.

To our participants' parents, without your participation in this study, this thesis and research would not be possible. Thank you for the opportunity to watch your children grow and most of all, your dedication and willingness to help facilitate infant biometric research.

I would also like to thank my parents. Your emotional support, encouragement, and unwavering faith in me, throughout graduate school and my life, has developed me into the person I am today. Specifically, to my father, Gary Hutchison, without the values of hard work, perseverance, and integrity that you have instilled in me, my successes in life would not have been possible. Every achievement I have had in life was because of you and the values you have taught me. To my beloved stepmother, Donna Hutchison, I would like to thank you for teaching me the value and importance of appreciation. You have been there for me all my life, you have loved and cared for me like a mother should. Your love has been unconditional and supportive, your advice

has been wise and if I’m being honest, basically always right. With that said, I will never forget the month I ate a ham sandwich or peanut butter and jelly for every meal because I refused to eat a meal you cooked. In hindsight, you have taught me an important lesson, cherish the things that others do for you because you never know when they may just stop. I appreciate all that you have done and continue to do for me and I am glad that no matter what you have always been there for me.

To my sister, Tasha Hull, I am so proud of everything that you have accomplished. You have always been one of the biggest role models in my life. You work hard, you do the right thing, and most of all you are a great person. Without your love, support, and guidance I would not be where I am at today.

Finally, to my future wife, Katrina Owens, the recently licensed R.N. – Through the good times and bad you have been there for me. There have been many long days and nights, of which you have supported me. You have been my rock throughout graduate school, without you and your support, following my passion and heart would not be possible. You have kept me fed when I forgot to eat, kept me sane when I was frustrated, and listened when I rambled. Thank you for allowing me to bounce ideas off you, reading and editing my work, and most importantly being there for me every step of the way. I can officially say that we did it, and I couldn’t be happier to start our “real” lives together.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Author: Hutchison, Torrey, L. MS
Institution: Purdue University
Degree Received: August 2018
Title: A Longitudinal Analysis on the Feasibility of Iris Recognition Performance for Infants 0-2 Years Old
Committee Chair: Stephen Elliott

The focus of this study was to longitudinally evaluate iris recognition for infants between the ages of 0 to 2 years old. Image quality metrics of infant and adult irises acquired on the same iris camera were compared. Matching performance was evaluated for four groups, infants 0 to 6 months, 7 to 12 months, 13 to 24 months, and adults. A mixed linear regression model was used to determine if infants' genuine similarity scores changed over time. This study found that image quality metrics were different between infants and adults but in the older group, (13 to 24 months old) the image quality metric scores were more likely to be similar to adults. Infants 0 to 6 months old had worse performance at an FMR of 0.01% than infants 7 to 12 months, 13 to 24 months, and adults.

CHAPTER 1. INTRODUCTION

There are only two existing studies (at the time of writing) that examined infant iris recognition (Bachenheimer, 2016; Corby et al., 2006). Corby et al. (2006) was the only study that examined iris recognition accuracy, and both studies reported capture rates and image quality findings. The studies showed that it was difficult capturing irises from infants younger than three, but when the iris was acquired, the iris had good quality. Comparing the quality results from the studies was difficult because both defined image quality differently, interpreting what a "good" or "low" quality sample is versus an "acceptable," "marginal," or "unacceptable."

Bachenheimer (2016) and Corby et al. (2006) used different iris recognition devices: the Iritech Binoculars and Panasonic Authenticam, respectively. The Iritech Binoculars are a portable, low-cost device which was priced at $480 (Fulcrum Biometrics LLC, 2018). The Panasonic Authenticam has been discontinued and was listed at $24.99 (eBay, 2018). Both studies attributed the low capture rates to the lack of cooperation from the younger infants, Bachenheimer mentioned that a more usable device may have improved the capture rate. Infants' lack of cooperation could degrade image quality and subsequently be rejected.

Jain et al. (2004) expressed the iris pattern does not stabilize until after the first two years of life and collecting irises from infants is quite difficult and therefore not feasible as a usable biometric trait for infants. This has been cited by many as evidence of the challenges of iris recognition with infants (Barra, Casanova, De Marsico, & Riccio, 2014; Barra et al., 2014; Jain et al., 2004; Jia et al., 2012; Tiwari et al., 2015; Tiwari, Singh, & Singh, 2013; Tiwari et al., 2013; Tiwari & Singh, 2012; Weingaertner, Bellon, Silva, & Cat, 2008).

Anatomically, the iris is known, in part from the ophthalmology community's research, to begin forming 6 weeks into gestation. The collagen fibrils, responsible for the pattern observed in the iris and used for biometric recognition (U.S. Patent No. 4,641,349, 1987), are fully formed before the end of the 7th month of gestation (Oyster, 1999; Remington, 2005). The formation of the collagen fibrils before birth make, at least theoretically, iris recognition performance feasible for infants directly after birth.

1.1 Statement of the Problem

Is iris recognition for infants between the ages 0-2 feasible?

1.2 Significance of the Problem

Infant identification has been studied on many biometric modalities such as footprint (Jia et al., 2012; Jia, Hu, Gui, & Lei, 2010; Kotzerke, Arakala, Davis, Horadam, & McVernon, 2014; Kotzerke, Davis, Horadam, & McVernon, 2013; Weingaertner et al., 2008), face (Bachenheimer, 2016; Bharadwaj, Bhatt, Singh, Vatsa, & Singh, 2010; Tiwari & Singh, 2012), fingerprint (Dutch Ministry of the Interior and Kingdom Relations, 2005; Jain et al., 2015, 2016; Jain, Arora, Cao, Best-Rowden, & Bhatnagar, 2017; Jain, Cao, & Arora, 2014), ear (Tiwari et al., 2015; Tiwari, Singh, & Singh, 2011, 2012c), and iris (Bachenheimer, 2016; Corby et al., 2006). There is an interest for infant biometrics for public health reasons such as biometric vaccination coverage which could potentially replace current methods such as identification cards, birth certificates, or where no identification method currently exists (Global Good Intellectual Ventures, 2017). Biometrics can also be used to find kidnapped or exploited children and thus it is important to find a biometric that remains stable over an individual's lifespan (Cole, 2016). Most infant biometrics such as footprint, face, fingerprint, and ear exhibit true accept rates of 70%-90% (Bharadwaj et al., 2010; Dutch Ministry of the Interior and Kingdom Relations, 2005; Jain et al., 2015, 2016, 2017, 2014; Jia et al., 2012, 2010; Kotzerke et al., 2014, 2013; Tiwari et al., 2015, 2011, 2012c, 2012c; Weingaertner et al., 2008). However, most of the studies that examined the performance of these biometrics capture samples only days apart making it difficult to determine if the biometric is feasible for long-term use (Bharadwaj et al., 2010; Jain et al., 2015, 2016, Jia et al., 2012, 2010; Lemes, Bellon, Silva, & Jain, 2011; Tiwari et al., 2015; Tiwari & Singh, 2012; Weingaertner et al., 2008).

Many studies dismiss the use of iris recognition based on a passage in Jain et al. (2004) which stated the iris pattern stabilizes sometime within the first two years of life (Barra et al., 2014, 2014; Jain et al., 2004; Jia et al., 2012; Tiwari et al., 2015, 2013, 2013; Tiwari & Singh, 2012; Weingaertner et al., 2008). To provide context, Jain's comment suggests that an iris captured from a one-month old infant may have a different iris pattern than a sample collected from the same iris and infant at two years old, which would mean that infant iris recognition over time is not stable.

There are many studies that also subsequently cite Jain et al, (2004), stating iris recognition for infants is difficult because collecting irises from infants is extremely difficult (Barra et al., 2014, 2014; Jain et al., 2004; Jia et al., 2012; Tiwari et al., 2015, 2013, 2013; Tiwari & Singh, 2012; Weingaertner et al., 2008). Though performance results reported by Corby et al. (2006), who correctly identified over 99% of infants between the ages of 1.5 to 8 years old a year after enrollment indicated that iris recognition for infants may be feasible. Neither, Bachenheimer (2016) or Corby et al. (2006), mentioned a change in iris patterns but they did have some difficulties capturing infant irises. In general, difficulty collecting biometric samples was a common theme that was highlighted throughout the infant biometric literature.

Additionally, Bachenheimer (2016) and Corby et al. (2006) used different iris recognition systems, with different image quality metrics, and different successful capture criteria. This study analyzed iris recognition performance using the same matching and quality assessment algorithm.

This study was unique, it was the first infant iris recognition study that:

- specifically examined infants 0 to 24 months old;
- compared image quality metrics and matching performance of adults and infants; and
- evaluated performance of different age groups.

1.3 Scope

This study examined if iris recognition performance for infants was feasible by examining performance at different age groups, performance over time, and quality metrics. A commercially available algorithm was used to assess quality, extract templates, and match templates. This is a secondary data analysis, as the data in this study came from a pre-existing dataset collected longitudinally.

1.4 Research Questions

This study examined the following research questions:

- Is there a difference between image quality metrics scores for adults and infants?
- Is there a difference in matching performance for different age groups?

- Do genuine similarity scores change over elapsed time?

1.5 Assumptions

The assumptions for this study included:

- All participants in this study were 0-2 years old;
- The image quality assessment algorithm and performance algorithm operated the same on infant irises as adult irises;
- All templates used in this study were large, a setting in the commercially available matcher;
- The matching algorithm operated on its slowest setting;
- The iris camera worked the same for infants as they do for adults; and
- The infants did not have any eye diseases.

1.6 Limitations

The limitations for this study included:

- The participants of this study may have been non-cooperative users; and
- This is a secondary data analysis.

1.7 Delimitations

The delimitations of this study included:

- Examining impostor similarity scores individually was outside the scope of this study;
- Examining subjects older than 24 months were beyond the scope of this study;
- Iris recognition algorithms and devices are designed for adults, which could impact how infant irises perform if they have different salient features than adult irises. This effect was beyond the scope of this study; and
- Examining infant behavior was beyond the scope of this study.

1.8 Definitions of Key Terms

Acceptable biometric capture attempt: “A capture attempt that fulfills the requirements of a biometric capture process” (JTC 1/SC 37, 2017, p. 12).

Biometrics: “Automated recognition of individuals based on their biological and behavioral characteristics” (JTC 1/SC 37, 2017, p. 2).

Biometric acquisition process: “Biometric capture process and additional processing to attempt to produce a suitable biometric sample(s) in accordance with the defined policy” (JTC 1/SC 37, 2017, p. 9).

Biometric permanence: “A biometric trait is permanent if it does not change over the lifetime of an individual” (Jain, Ross, & Nandakumar, 2011, p. 13).

Biometric sample: “An analog or digital representation of a biometric characteristics prior to biometric feature extraction” (JTC 1/SC 37, 2017, p. 6).

Captured biometric sample: “A biometric sample resulting from a biometric capture process” (JTC 1/SC 37, 2017, p. 7).

Character: “Contributor to quality if a sample attributable to inherent features of the source” (JTC 1/SC 37, 2012, p. 2).

Child: “A person 6 to 12 years of age. An individual 2 to 5 years old is a preschool child.” (Online Medical Dictionary, 2018a).

False match rate: “Proportion of the completed biometric non-match trials that result in a false match” (JTC 1/SC 37, 2012, p. 3).

False non-match rate: “Proportion of the completed biometric match trials that result in a false non-match” (JTC 1/SC 37, 2012, p. 3).

Fidelity: “Expression of how accurately a biometric sample represents its source biometric characteristic” (JTC 1/SC 37, 2012, p. 2).

Gestation: “The period of development in the uterus from conception until birth” (Farlex, 2018)

Gray scale utilization: “Measures the overall iris image for evidence of a spread of intensity values in iris data” (JTC 1/SC 37, 2011, p. 6).

Ground truth: “A set of data that is considered to be accurate and reliable, and is used to calibrate a model, algorithm, procedure, etc.” (Oxford Dictionary, 2018b).

Infant: “A child between 1 and 23 months of age.” (Online Medical Dictionary, 2018b).

Iris boundary shape: "A mathematical expression of the iris sclera boundary and its deviation from circularity" (JTC 1/SC 37, 2011, p. 8).

Iris-pupil boundary contrast: "Represents the image characteristics at the boundary between the iris region and the pupil" (JTC 1/SC 37, 2011, p. 9).

Iris pupil concentricity: "The degree to which the pupil center and the iris center are in the same location" (JTC 1/SC 37, 2011, p. 10).

Iris radius: "Represents the distance across the iris along the horizontal" (JTC 1/SC 37, 2011, p. 11).

Iris-sclera boundary contrast: "Represents the image characteristics at the boundary between the iris region and the sclera" (JTC 1/SC 37, 2011, p. 10).

Quality: "Quantitative value of the fitness of a biometric sample to accomplish or fulfil the comparison decision" (JTC 1/SC 37, 2017, p. 21).

Margin: "The degree to which the image achieves positioning of the iris portion of the image relative to the edges of the entire image" (JTC 1/SC 37, 2011, p. 12).

Principal component analysis: "A method of analysis which involves finding the linear combination of a set of variables that has maximum variance and removing its effect, repeating this effectively (Oxford Dictionary, 2018a).

Pupil boundary shape: "A mathematical expression of the iris pupil boundary and its deviation from circularity" (JTC 1/SC 37, 2011, p. 13).

Pupil to iris ratio: "The degree to which the pupil is dilated or constricted... the pupil to iris radius (JTC 1/SC 37, 2011, pp. 13–14).

Sameness: "Whether image pairs with similar quality values give lower FNMR than others" (Tabassi, Grother, & Salamon, 2011, p. 10).

Sharpness: "The degree of defocus present in the image" (JTC 1/SC 37, 2011, p. 14).

Soft biometrics: Soft biometrics are biometric traits that are not unique across the population e.g., height, weight, eye color.

Stability: The change of performance with regards to a specified covariate (O'Connor, 2013).

Usable iris area: "The percent of the iris portion of the image that is not occluded by eyelids, eyelashes, or saturating specular reflections, expressed as the percentage of area of an annulus modeling the iris without such occlusions" (JTC 1/SC 37, 2011, pp. 14–15).

Utility: “The observed performance of a biometric sample or set of samples in one or more biometric systems. The character of the sample source and the fidelity of the processed samples contribute to – or detract from – the utility of the sample” (JTC 1/SC 37, 2012, p. 4).

CHAPTER 2. LITERATURE REVIEW

This chapter provides a review of literature that covers the following general topics: biometrics, biometric performance, biometric image quality, infant biometrics and their challenges, the structure and development of the eye, iris recognition, and iris aging. The literature review was used to identify gaps in the literature and build a methodology that examined the feasibility of infant iris recognition.

2.1 Biometrics

Biometric recognition uses characteristics, behavioral or biological, to identify or verify an individual's identity. For a biometric characteristic to be useful it must possess the following traits: remain similar throughout an individual's life time (i.e. permanence), be a common characteristic of a population (i.e. universal), differ from individual to individual (i.e. uniqueness), suitable matching rates for a specified application (i.e. performance), easy to collect and measurable (i.e. measurability), generally accepted by the population (i.e. acceptability), and difficult to fake or alter, i.e. circumvention (Clarke, 1994).

A biometric system consists of subsystems that are present in most biometric systems and therefore can be generalized to fit a generic model for biometric systems. The general biometric model was created to explain the functions of a biometric system (Mansfield & Wayman, 2002). The subsystems of the general biometric model comprise of data capture, signal processing, data storage, matching, and decision making.

The data capture subsystem captures a raw biometric sample from a user's presentation to the biometric sensor and sends the sample, as a signal, to the signal processing subsystem (Mansfield & Wayman, 2002). The signal processing system decides to reject or accept an image base off of a pre-set quality threshold (JTC 1/ SC 37, 2006; Mansfield & Wayman, 2002). If a sample is rejected, the biometric system may attempt to capture another biometric sample; otherwise, the extracted features are stored as a template in the data storage subsystem or used directly by the matching subsystem.

Templates are created and stored during enrollment. The extracted features are not stored directly during matching. Templates from the data storage subsystem are used in the matching subsystem to generate similarity (or dissimilarity) scores to be used to verify or identify the user who has presented to the biometric system (Mansfield & Wayman, 2002).

Samples are acquired at the data capture subsystem and can be affected by an individual's interaction with the device e.g., a non-cooperative subject. For example, a non-cooperative subject may look away from the iris camera making it difficult to capture an iris sample. Poor quality samples, without quality control, could propagate throughout the whole system making it difficult to correctly extract biometric features, match, and result in a false rejection or acceptance (Wayman, 2000). Therefore, it is important that infants' iris image quality was examined, indicating whether samples of good quality can be given repeatedly.

2.2 Performance

When a sample is collected, a comparison is attempted against an enrolled template. If the collected sample and enrolled template share the same ground-truthed identity, then it is considered a genuine match. Conversely, if they do not share a ground-truthed identity then it is considered an impostor match. In an operational setting, an individual's identity cannot be ground-truthed. When a biometric system performs matching it computes a similarity score that determines whether a user is accepted or rejected based off a predetermined threshold value. The decision subsystem of a biometric system determines a binary classification "yes" if the similarity score is above or equal to the threshold and "no" if the similarity score falls below the threshold (Mansfield & Wayman, 2002). The classification by the biometric system can result in four outcomes: an impostor match is correctly rejected, an impostor match is falsely accepted, a genuine match is correctly accepted, and a genuine match is falsely rejected. A false accept is analogous to a security breach and a false reject results in an inconvenience to the user.

If several matches have been conducted, a score histogram is created to plot the impostor and genuine scores. Figure 2.1 is an example of a score histogram with a genuine (blue) and impostor (red) distribution. The horizontal line represents a set arbitrary threshold of 50. Any match score at or above 50 is accepted into the system and any match score below is rejected. The farther the impostor and genuine score distributions are from each other the better the system is at discriminating genuine users from impostors. Moreover, if no overlap between the genuine and impostor distributions exists, then a threshold value can be chosen that results in no false accepts or rejects.

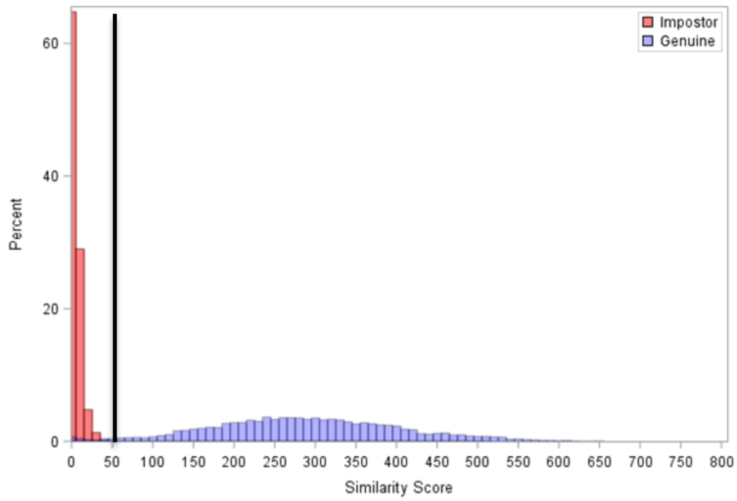

Figure 2.1. Similarity Score Histogram

The false match rate (FMR) is the proportion of impostor attempts that are greater than or equal to the threshold, where the false non-match rate (FNMR) represents the proportion of genuine attempts that are below the threshold (Dunstone & Yager, 2009; JTC 1/ SC 37, 2006). Equations 1 and 2, show the false match rate and false non-match rates for a given threshold *t*, respectively.

$$FMR = \frac{\#\ of\ impostor\ attempts \geq t}{Total\ \#\ of\ impostor\ attempts} \tag{1}$$

$$FNMR = \frac{\#\ of\ genuine\ attempts < t}{Total\ \#\ of\ genuine\ attempts} \tag{2}$$

A failure-to-acquire (FTA) occurs when a system fails to capture a biometric sample. An acquisition can fail because the biometric characteristic could not be presented; a sample cannot be segmented; a sample's features cannot be extracted; or a sample's extracted features do not meet quality control thresholds. A FTA is the proportion of attempts the biometric system failed to capture a sample (JTC 1/ SC 37, 2006). The false reject rate (FRR) is the proportion of genuine transactions that were rejected by the system and the false accept rate (FAR) is the proportion of impostor transactions that were falsely accepted by the system (Dunstone & Yager, 2009; JTC 1/ SC 37, 2006). Equations 3 and 4 show the false reject rate and the false accept rate, respectively.

The FRR and FAR account for the genuine and impostor attempts and examine the number of images that failed to be acquired.

$$FRR=FTA+FNMR*(1-FTA) \quad (3)$$

$$FAR=FMR*(1-FTA) \quad (4)$$

There are two fundamental types of performance - verification and identification. Verification occurs when a person makes a claim to an identity and the captured biometric is compared to the template stored under the identity claimed. Identification occurs when all the templates stored in a database are compared to a captured biometric, returning a list of potential candidates. The number of potential candidates is pre-determined and is primarily denoted as a rank e.g., rank-1 identification returns the highest similarity score (JTC 1/ SC 37, 2006).

To evaluate biometric performance across all thresholds, a Detection Error Trade-off (DET) curve is used. A DET curve is a modified receiver operating characteristic curve (ROC) that plots the FNMR (or FRR) against the FMR (or FAR) of a biometric system (Dunstone & Yager, 2009; JTC 1/ SC 37, 2006). The DET curve represents the trade-off between FMR (or FAR) and FNMR (or FRR) as the threshold is varied. A higher threshold results in a lower FMR (or FAR) and a higher FNMR (or FRR) and vice versa for a lower threshold (Dunstone & Yager, 2009). An ideal DET curve will have a 0% false match rate for all possible false non-match rates, and a 0% false non-match rate for all possible false match rates. Graphically, the curve would lie directly on the x-axis and y-axis.

Studies show that iris recognition performance can be affected by blurriness or severely occluded irises; these defective iris images can be detected using image quality assessment tools or visual investigation from a test administrator. Performance of an iris recognition algorithm will vary based on an algorithm's specific sensitivity to certain characteristics of an iris image. Algorithms may be more robust or sensitive to severely constricted and dilated pupils, poorly centered irises, saturated images, specular reflections, and high grey level images. It has also been observed that dilation differences between mated pairs of iris images can increase the false non-match rate as well. Image quality is a useful quantitative measure that can be used to predict performance. Images of higher quality would expect to have a higher similarity score between mated pairs than an image of low quality (Grother et al., 2012).

2.3 Image Quality

It is important to discuss image quality, a metric meant to be a predictor of a biometric system or matcher's performance. This section outlines the definition of image quality and its use in biometrics. Furthermore, it outlines iris image quality metrics, their measurement, and definition in accordance to ISO 29794-6.

2.3.1 Introduction to Image Quality

Image quality is a quantitative value used for predicting a biometric matcher's performance e.g., a system with low quality images may have difficulty extracting features and would perform better with higher quality samples (Tabassi, Wilson, & Watson, 2004). Image quality can be used to reject low quality samples in favor of samples with higher quality, define quality thresholds for enrollment, and establish a higher weight for high quality samples in biometric fusion schemes (Maltoni, Maio, Jain, & Prabhakar, 2009; Tabassi et al., 2004). An image quality assessment algorithm can also be used to improve biometric samples by specifying a reason why a particular sample is poor and presenting corrective feedback to the user or operator.

Generally, there will only be a small number of low quality samples compared to high quality samples; the small proportion of low quality samples will still impact performance. Samples of low quality decrease the chances of a correct match and increase the number of false negatives. Samples of extremely low quality may not be able to attempt to verify or identify. Sensor and user interface design can improve the way a subject interacts or uses the device while simultaneously improving image quality. The illuminator and optics can also be improved to collect higher quality samples; restricting the environment and other confounding variables will increase consistency across collected samples. Better samples can also be collected by adhering to data collection best practices (Tabassi et al., 2011).

There are several image properties and iris characteristics that influence performance of iris recognition. The quality of an image can be determined with an overall scalar quality score or can be broken down into more detailed image quality metrics that represent various aspects of the iris image known to influence performance. The scalar quality score is used to identify poor quality samples and exclude them by setting a quality threshold, and image quality metrics give more information that is actionable for feedback to the user or operator (Tabassi et al., 2011).

Iris image quality metrics recorded by Neurotechnology 10 SDK include: scalar quality, usable iris area, iris pupil contrast, iris sclera contrast, pupil boundary circularity, iris pupil concentricity, sharpness, pupil-to-iris ratio, interlace, grayscale spread utilization, iris radius, margin adequacy, and iris detection confidence. The image quality metrics provided by Neurotechnology, except for iris detection confidence, adhere to the image quality data standards outlined in ISO/IEC 29794-6 (Neurotechnology, 2017).

2.3.2 Description of Iris Image Quality Metrics

Scalar quality scores should predict performance metrics. All image quality assessment algorithms should compute a score so that the false non-match rate will increase for low quality samples and decrease for those of higher quality (Tabassi et al., 2011). Therefore, the highest image quality score should produce lower error rates than lower scores (Tabassi et al., 2011).

The amount of the iris that is not occluded by specular reflections, eyelids, or eyelashes is referred to as the usable iris area. A lower usable iris area indicates that there is less information to extract from the iris image for recognition. The usable iris area is represented as the percentage of the iris area that is not occluded and is recommended to be at least 70% (JTC 1/SC 37, 2011; Tabassi et al., 2011). Subject behavior and the collection environment may impact the usable iris area. The usable iris area can be improved by designing a better iris recognition system that reduces specular reflection from the system's illuminator, employs automatic quality control, and improves subject interaction (Tabassi et al., 2011).

Iris pupil contrast is the degree of contrast at the boundary between the pupil and the iris. The higher degree of contrast between the iris and pupil the easier it is for an iris to be segmented; the contrast of the pupil and iris is less than the contrast between the iris and sclera. The degree of contrast between the pupil and iris will vary for each subject whether an image is captured in the visible light spectrum or the typical near-infrared spectrum. The contrast between the pupil and iris can also be affected by the iris recognition system; the level of contrast is dependent on the illuminator of the device (Tabassi et al., 2011). The measure of iris pupil contrast is dimensionless and is scored as the percentage of contrast between the pupil and the iris at the iris pupil boundary, a recommended iris pupil boundary score is 30% or higher (JTC 1/SC 37, 2011).

The pupil shape is the regularity of the pupil iris boundary. The shape of the pupil is not expected to be completely circular or even elliptical but is measured as the deviation of the pupil boundary from a circular shape (JTC 1/SC 37, 2011; Tabassi et al., 2011). The circularity of the pupil boundary is a function of subject behavior and inherent anatomy. The shape of the pupil boundary will vary person to person and a non-circular boundary can also be caused by a non-frontal gaze to the iris camera (Tabassi et al., 2011).

Iris sclera contrast is the degree of contrast at the boundary of the iris and sclera. The contrast between the iris and sclera is scored as a percentage of contrast between the iris and sclera at the boundary, the iris sclera contrast should be greater than 5% (JTC 1/SC 37, 2011; Tabassi et al., 2011). The iris sclera contrast varies for each person and is also dependent on illumination which is affected by the iris recognition system, surrounding environment, or both. The contrast between the iris and sclera can be improved by designing a better acquisition and capture process (Tabassi et al., 2011).

This iris pupil concentricity measures the degree that the pupil and iris share the same center. The center of the iris and pupil may not be the same and large deviations from concentricity can cause segmentation errors. The concentricity of the pupil and iris is measured by taking the distance between pupil and iris centers and dividing by the radius of the iris. The iris pupil concentricity should be less than a fifteenth of iris's radius (JTC 1/SC 37, 2011).

Sharpness measures the absence of defocus blur in an image. An object outside of a camera's depth of field would cause defocus blur and would become more pronounced as an object moves further away from the focal plane. The impairments caused by defocus blur is like motion blur; therefore, blur deficiencies caused by motion may be detected by sharpness. Camera characteristics such as the aperture size can affect the depth of field. Also, the user interface could be improved to guide the subject to a proper distance to reduce the chance of the iris being outside of the focal plan of the iris recognition system (Tabassi et al., 2011).

Pupil-to-iris ratio measures the degree of dilation by dividing the pupil radius by the iris radius. The recommended pupil to iris ratio is between 0.2 and 0.6 assuming that the average of an iris radius is 6 millimeters (JTC 1/SC 37, 2011; Tabassi et al., 2011). Iris recognition performance will tend to degrade for extreme values of pupil to iris ratio. Dilation depends on the subject's behavior e.g., drugs or ambient light from the environment the iris images are acquired from (Tabassi et al., 2011).

The dilation change is the difference between two iris samples, measured with pupil-to-iris ratios and accounting for magnification effects. The dilation change ΔD, as shown in Equation 5, is the ratio of the two iris radii, D_1 and D_2. Dilation change assumes that the iris remains a constant anatomical size and that the pupil-to-iris ratio D_1, is greater than D_2.

$$\Delta D = \frac{R_{I1}}{R_{I2}}\left(\frac{R_{I2}-R_{P2}}{R_{I1}-R_{P1}}\right) = 1 - \frac{1-D_1}{1-D_2} \tag{5}$$

The gray scale utilization is the degree in which an image is exposed to a wide range and distribution of intensity values of pixels. Underexposed images have few high intensity pixels which results in a darker, more blackish, image. Over exposed images have few low intensity pixels which results in a saturated, more whitish, image. Poor illumination or over saturation of an image can cause a small spread of intensity values. Gray scale utilization is measured in bits and is a result of the entropy obtained from an image's pixel histogram. The higher the entropy the more exposed an image is. The gray scale utilization of an iris image should at least be 6 (JTC 1/SC 37, 2011; Tabassi et al., 2011). Correcting an iris recognition system to produce images of higher contrast and dynamic range can improve gray scale utilization. Gray scale utilization is also impacted by the environment in which images are acquired (Tabassi et al., 2011).

The iris radius is measured by the number of the pixels across the radius of the iris. An iris should be at least 60 pixels across (JTC 1/SC 37, 2011). The iris radius can be affected by the sampling rate of the image acquisition device or the distance a subject is from the device. The iris radius can be improved by better positioning a user to the iris acquisition device.

Margin adequacy is the degree an iris is from the boundary of the image. Inadequate iris margin differentials occur from incorrect segmentation of the iris which can be caused by subject movement at the time of capture. A margin adequacy score of 100 indicates that the margin values are at least the margin values specified in ISO/IEC 19794-6:2011 (JTC 1/SC 37, 2011). Improving user interaction with a better user interface can improve margin adequacy. The underlying segmentation algorithm may also need changed to improve segmentation (Tabassi et al., 2011).

Interlacing artifacts are caused by misaligned odd and even rows of pixels and can result in loss of vertical resolution. Interlacing is typically an issue seen in legacy cameras and is a direct effect of the device used to acquire irises (Tabassi et al., 2011). Iris samples collected independently from separate occasions for the same individual can differ from a change in acquisition environment (e.g., illumination), subject's presentation to the system (e.g., behavior, habituation), physical changes of the biometric (e.g., pupil size, occlusion, disease, etc.), changes to the sensor itself (e.g., sensor aging, different sensors). These variations can influence the similarity score of an individual resulting in worse false non-match rates (Grother, Matey, Tabassi, Quinn, & Chumakov, 2013; Tabassi et al., 2011).

Table 2.1 summarizes image quality metrics and their effect on false non-match rates and if the pairwise quality1 also changed performance. Moreover, it indicates if an image quality metric is affected by behavior, environmental conditions, device specific characteristic, or natural anatomical variation (excluding diseases or defects) it is coded with a yes in Table 2.1.

[1] Pairwise quality is calculated with the geometric mean of two samples from the same individual e.g., $\sqrt{q_1 * q_2}$

Table 2.1.
Image quality metrics and its causes and effects on FNMR. This table was adapted from (Grother et al., 2012; Tabassi et al., 2011)

Quality Metric	Does it affect FNMR	Does Sameness Matter	Source of Impairment	Subject Behavior	Subject Character (natural anatomical variation)	Environment	Device
Scalar Quality	Yes	No	-	-	-	-	-
Usable Iris Area	Yes	Yes	Occlusion (eyelids, eyelashes, specular reflections)	Yes	-	Yes	Yes
Iris Pupil Contrast	Yes	Yes	Intrinsic, Illumination	-	Yes	Yes	Yes
Pupil Boundary Circularity	Yes	No	Disease, Off Axis Disease	Yes	Yes	-	-
Iris Sclera Contrast	Yes	Yes	Intrinsic, Illumination	-	Yes	Yes	Yes
Sharpness	Yes	No	Defocus, Compression	Yes	-	-	Yes
Dilation	Yes	Yes	Ambient light, Intrinsic	Yes	Yes	Yes	Yes
Interlace	Yes	No	Loss of vertical resolution	-	-	-	Yes
Gray Scale Spread	Yes	Yes[2]	Illumination, Saturation	-	-	Yes	Yes
Iris Radius	Yes	Yes	Resolution, Distance to Camera	Yes	-	-	Yes
Margin	Yes	No	Improper Crop, Subject-device Alignment	Yes	-	-	Yes

[2] The impact varied based on the algorithm

Scalar quality is a quantitative indicator of performance (Tabassi et al., 2004). If an image has high scalar quality, then a lower FNMR would be experienced compared to images that have low quality. Image quality metrics can affect performance because of several factors such as: subject behavior, collection environment, device characteristics, and natural anatomical variation. Image quality metrics can affect performance for extremely low or high values (e.g., pupil to iris ratio) or if the pairwise quality of the samples being matched differ (e.g., usable iris area). It is also important to note that different performance and image quality assessment algorithms can be more robust or sensitive to certain image quality metrics. In summary, an assessment of genuine similarity scores or false non-match rates should consider individual image quality metrics and if they stay the same between different samples from the same individual.

2.4 Infant Biometric Performance

2.4.1 Footprint Recognition

Two studies, Jia et al. (2010) and Jia et al. (2012), tested the biometric performance of several algorithms for infant footprint recognition. Footprint samples were captured during one session within the first two days following birth, approximately 19-20 samples were collected from each infant (Jia et al., 2012, 2010).

Jia et al. (2010) examined an algorithm with three different similarity score measures. The best identification rate was 97% with an EER of 3.82%, a false accept rate was not disclosed. Jia et al. (2012) examined the performance, verification rate, and four different footprint algorithms. The best performing footprint algorithm had a verification of 98% with a FAR of 0.001%. Both studies stated collecting footprint samples from newborns was difficult because they were extremely irritated due to hunger or tiredness and would cry often. When the infants were sleeping or calm acquiring images was much easier than when they were upset (Jia et al., 2012).

Weingaertner et al. (2008) attempted collecting footprints and palmprints using the traditional ink and paper method, optical fingerprint/palmprint scanners, and high-resolution light scanner. The ink and paper method did not provide much information, the footprint samples did not have many visible ridges rendering the inked prints unsuitable for identification. The optical fingerprint/palmprint scanners, 250dpi to 500dpi, also lacked usable ridge patterns making the prints unsuitable for recognition. Additionally, a high-resolution light scanner was tested at

1200dpi and 2400dpi, resulting in higher quality samples. However, the infants' feet/palms had to be held still for approximately two minutes or else the images would get distorted, and the contrast between the ridges and valleys was low, making it difficult to segment features. Due to the failure of the other devices, a 1400dpi sensor was developed using an 8-megapixel camera that was attached to an optical glass prism. Using this sensor, two prints were collected – within the first 24 hours and second 24 hours after birth – from each infant's palm and foot. The best palmprint and footprint sample was taken from the first 24-hour visit and classified into five different quality categories. A quality rating of excellent was rewarded when the ridge pattern, deltas, and minutiae points were clearly visible, and a good quality rating when the ridge pattern and delta(s) were visible, but minutiae points were not. Only 37.7% of the infants had good or excellent quality rating and were not sufficient to attempt matching.

Kotzerke et al. (2013) created an algorithm that extracts the flexure creases on the bottom of the foot instead of the ridge patterns for infant verification. The flexure creases are represented by the darker lines on the bottom of the foot in Figure 2.2. Fifty-four sets of footprints were collected at 0-3 days, 8 weeks, and 6 months old. After flexure creases were extracted, matching was performed manually with 20 footprint pairs, 11 from the same infant and 9 from different infants. Seven individuals, classified as non-experts in biometrics, correctly verified the infants 55% of the time, two ride-based biometric experts correctly verified 95% of the flexure crease pairs. The algorithm was able extract flexure creases but there was a trade-off for the optimal contrast threshold. If the contrast is set too low than some creases cannot be extracted and creates false creases if the contrast threshold is set too high.

Using the same data set from Kotzerke et al. (2013), the area under the big toe (i.e. the ball print) was used to identify infants, which is a ridge-based biometric that uses minutiae points like fingerprint recognition – Figure 2.2 outlines the ball of the foot with a black box (Kotzerke et al., 2014). Neurotechnology's VeriFinger software development kit (SDK), a commercially available fingerprint software, was used to extract and match minutiae from infants' ball prints. The ball prints were collected with the NEC PU900-10, a commercial fingerprint sensor. Ball prints collected during the first visit were low quality and excluded from the performance analysis. The intra- and inter-visit performance was examined for visit 2 and visit 3, and inter-visit performance was analyzed by treating visit 3 as the stored image and visit 2 as the matching image. In an operational scenario, visit 2 would be the stored image and would yield worse equal error rates from the lower quality samples collected at a younger age. Intra-visit performance was calculated for visit 2 and visit 3 and inter-visit performance between visit 2 and visit 3.

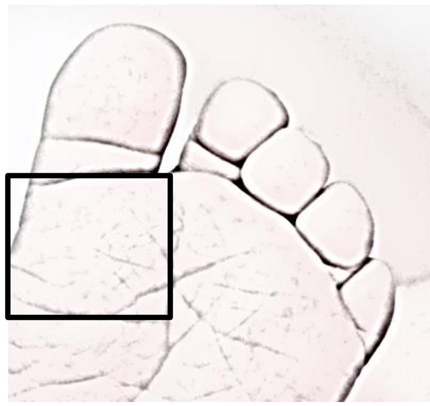

Figure 2.2. Ball of the Foot. The ball of the foot is highlighted in black. This image was modified from the original image (Pexels, 2018).

The intra-visit performance, using the Neurotechnology VeriFinger SDK, produced an EER of 16.60% and 14.28% for visit 2 and 3, respectively and EER of 29.34% for inter-visit performance. The infant's ball print ridges were smaller than a typical adult's fingerprint ridges, therefore the resolution was reduced to accommodate the difference in ridge sizes. Reducing the resolution improved the intra-visit performance EER's to 0% for visit 2 and visit 3 and 7.28% for inter-visit performance. Two additional matching algorithms were used to perform matching with minutiae data that was extracted by Neurotechnology VeriFinger SDK. Both algorithms did not perform as well as the matching algorithm implemented by Neurotechnology.

2.4.1.1 <u>Performance Summary</u>

Infant footprint recognition performance was affected by infants becoming agitated and crying. It was easier to collect samples when the infants were calm or upset. One device took around two minutes to capture a sample which may increase the chance in infant becomes agitated. Table 2.2 summarizes the performance results for each foot-based biometric infant study and specifies the part of the foot used. The footprint algorithm implementation is also listed and is followed by the FAR, EER, performance type (e.g., identification or verification), and the quality metrics reported. To simplify comparisons the number of visits and the respective age in each visit is denoted inside the parentheses e.g., 2 (first 24 hours, second 24 hours).

Table 2.2.
Performance summary results for infant foot-based biometrics

Article	Modality	Algorithm	Acceptance Rate/Performance				EER	Type	Quality	# Of Visits (age(s))
			FAR = 0.001%	FAR = 0.01%	FAR = 0.1%	Unspecified				
Jia et al., 2010	Footprint	BLPOC w/ Peak	-	-	-	95.05%	4.34%	Acceptance Rate (Identification), EER (Verification)	-	1
		BLPOC w/ PCE	-	-	-	93.30%	4.52%		-	
		BLPOC w/ PSR	-	-	-	97%	3.82%		-	
Jia et al., 2012	Footprint	Ordinal Code	96.2%	96.6%	97.2%	-	1.5	Verification	-	1
		Competitive Code	95%	95.8%	96.2%	-	2.2		-	
		BOCV	98%	98.2%	98.5%	-	1.34		-	
		RLOC	96.8%	97.5%	98%	-	1.77		-	
Weingaertner et al., 2008	Footprint	n/a	-	-	-	-	-	-	37% excellent or good quality	2 (1st 24hrs, 2nd 24hrs)
Kotzerke et al., 2013	Flexure Creases	Manual Inspection	-	-	-	55% (Non-expert) 95% (Expert)	-	Verification	-	3 (0-3 days, 8 weeks, 6 months)
Kotzerke et al., 2014	Ballprint	Verifinger	-	-	-	-	16.6% (V2, V2) 14.28% (V3, V3) 29.34% (V3, V2)	Verification	-	3 (0-3 days, 8 weeks, 6 months
		Verifinger (re-scaled resolution)	-	-	-	-	0% (V2, V2) 0 (V3, V3) 7.28% (V3, V2)		-	
		ICP	-	-	-	-	45.75% (V2, V2) 40.72% (V3, V3) 44.9% (V3, V2)		-	
		BGM	-	-	-	-	14.66% (V2, V2) 16.01% (V3, V3) 40.08% (V3, V2)		-	

2.4.2 Palmprint Recognition

Weingaertner et al. (2008) also collected two palmprints over two separate visits, the first 24 hours after birth and the second 24 hours after birth. The best palmprint sample from each infant was classified into 5 distinct levels of quality, excellent and good quality classifications are defined in Section 2.4.1, and 83% of the collected palmprint samples had a quality classification of excellent or good. Three fingerprint examination experts manually matched two sets consisting of 30 randomly selected infant palmprint pairs. The fingerprint experts correctly verified 63.3% and 67.7% of the first and second set of the palmprint pairs, respectively.

Another palmprint data collection occurred at the same hospital as Weingaertner et al. (2008) and collected 1,221 samples from 250 newborns between 1-48 hours after birth. Five sets containing three samples from each infant's right palm were collected using the Crossmatch LSCAN 1000P, a commercially available fingerprint/palmprint scanner. Palmprint quality was assessed automatically from classifications methods proposed in Wu, Tulyakov, and Govindaraju, (2006), which include good, normal, dry, wet, and spoiled. Moreover, good quality is defined by the traits "clear ridge/valley contrast; easily-detected ridges; precisely-located minutiae; easily segmented" (Wu et al., 2006, p. 217). Of the infants, only 5% (i.e. 20 out of 250 newborns) had good quality palmprint samples, consequently, the same proportion of infants' samples were sufficient for matching. Many of the palmprint samples did not have visible ridge structures, minutiae points, or deltas making it difficult to perform matching on most of the images. The best performing algorithm, simulated annealing (SA), had a verification rate of 78% at a FAR of 1%. Moreover, the rank-3 identification rate was 98% and identification rates at ranks larger than three were 100% (Rhcastilhos, 2018).

2.4.2.1 Performance Summary

The performance of palmprint recognition for infants was increased from a manual verification rate of 63.33-67% to 78% with the SA matching algorithm. However, 5% of the infants had palmprints that had enough quality for matching, furthermore, manual assessment of image quality resulted in 83% of good or excellent quality palmprint images which may be influenced by the subjective quality measurements of the examiners. Table 2.3 conveys the performance and image quality results of each palmprint recognition study. Additionally, Table 2.3 reports the corresponding FAR, algorithm, and performance type (i.e. verification or identification). Comparisons between studies is simplified by reporting the number of visits and respective age for each study e.g., 2 (first 24 hours, second 24 hours).

Table 2.3.
Performance summary results for infant palmprint recognition

Article	Modality	Algorithm	Acceptance Rate/Performance				EER	Type	Quality	# Of Visits (age(s))
			FAR = 0.1%	FAR = 1%	FAR = 10%	Unspecified				
Weingaertner et al., 2008	Palmprint	Manual Inspection	-	-	-	63.33%	-	Verification	83% of good and excellent quality	2 (1st 24hrs, 2nd 24hrs)
			-	-	-					
			-	-	-	67%				
Lemes et al., 2011	Palmprint	SA	-	78%	-	-	-	Verification	5% good quality	1
			-	98%	-	-	-	Rank 3 Identification		

2.4.3 Face Recognition

Bharadwaj et al. (2010) collected face images from 34 newborn infants over two sessions, two hours after birth and again at the infant's discharge from the hospital. A face recognition algorithm that combines the scale and rotation invariant descriptors algorithm (SURF) and texture operator algorithm (LBP) was proposed. The proposed algorithm performance was compared to PCA, LDA, ICA, SURF, and LBP algorithms. In fact, the proposed algorithm had the best rank-1 identification rate 86.9%; in comparison, the worst performing algorithm was LBP with a rank-1 identification rate of 80.1%.

Another study using the same algorithms in Bharadwaj et al. (2010) collected face images in two sessions, first four hours after birth and again 20-70 hours after birth (Tiwari & Singh, 2012). Infants were crying, sleeping, or agitated making it difficult to capture face images with a neutral expression. Each algorithm was trained and tested with face images based on the classified expressions neutral, crying, and screaming. The proposed algorithm had the best rank-1 identification for all training and testing combinations. When neutral faces were used for training and testing the proposed algorithm had a rank-1 identification rate of 87.04% and outperformed the rank-1 identification rate observed by Bharadwaj et al. (2010). Moreover, training the algorithm with neutral faces and testing with sleeping or crying resulted in a higher rank-1 identification rate than when crying or sleeping expressions were used for training.

Bachenheimer (2016), using a low-cost and portable biometric system, observed that 57% of infants' face samples from ages 0-3 years old produced good quality images. Moreover, 42% of the face samples were of low quality and 1% failed to acquire an image at all. Image quality for infants four years and older increased the proportion of good quality samples to 79% and decreased the proportion of low quality and non-acquired samples to 20% and 1%, respectively.

2.4.3.1 Performance Summary

Table 2.4 summarizes the performance and image quality results from each infant face recognition study and face recognition algorithms. Moreover, the table denotes the testing-training combinations and the number of visits and respective age e.g., 2 (first two hours, at discharge).

Table 2.4.
Performance summary results for infant face recognition

Article	Algorithm	Rank-1 Identification Accuracy N-N[3]	N-C	N-S	C-N	S-N	C-S	Not Categorized	Quality	# of Visits (age)
Bharadwaj et al., 2010	PCA	-	-	-	-	-	-	81.3%	-	2 (First 2 hours, at discharge)
	LDA	-	-	-	-	-	-	80.7%	-	
	ICA	-	-	-	-	-	-	84.6%	-	
	LBP	-	-	-	-	-	-	82.4%	-	
	SURF	-	-	-	-	-	-	80.1%	-	
	Proposed	-	-	-	-	-	-	86.9%	-	
Tiwari & Singh, 2012	PCA	81.88%	74.58%	80.38%	75.58%	77.58%	68.38%	-	-	2 (First four hours, 20-70 hours after birth)
	LDA	83.19%	81.39%	85.39%	81.29%	81.39%	71.29%	-	-	
	ICA	83.34%	81.64%	84.14%	80.15%	81.35%	72.64%	-	-	
	LBP	83.84%	79.64%	81.54%	77.34%	80.34%	75.24%	-	-	
	SURF	82.16%	80.36%	81.46%	80.36%	80.36%	78.35%	-	-	
	Proposed	87.04%	85.14%	86.34%	84.34%	85.84%	81.34%	-	-	
Bachenheimer, 2016	n/a	-	-	-	-	-	-	-	Good (57%) Low (42%) Not Captured (1%)	1 (0-3yrs)
	n/a	-	-	-	-	-	-	-	Good (79%) Low (20%) Not Captured (1%)	1(4+ yrs)

[3] The notation refers to the training and testing set, training-testing. Neutral, sleeping, and crying facial expressions are abbreviated with N, S, and C.

2.4.4 Ear Recognition

Tiwari, Singh, and Singh (2011) initially investigated ear recognition for infants with four algorithms, collecting 5 samples of the left and right ear from 125 subjects. The highest performing algorithm, geometrical feature extraction (GF), had a rank-1 identification rate of 83.67%. In 2012b, Tiwari, Singh, and Singh compared the performance of infant and adult ear recognition. The infant database consisted of 210 subjects with 5 samples per ear, and the adult database had 121 different subjects, 14-58 years old, with 471 images total. Seven separate algorithms were also compared, the best performing algorithm for infants, HAAR, had a rank-1 identification rate of 91.23% and 93.5% for infants and adults, respectively. The lowest performing algorithm for infants, PCA, had a rank-1 identification accuracy of 81.14% and 83.32% for infants and adults, respectively. The results of this study indicate that ear recognition for infants and adults have similar rank-1 identification rates. However, the adult database did not have as many subjects as the infants and the time between samples was not given, potentially resulting in misconstrued conclusions.

In another study, infant ear recognition was tested using six different matching algorithms PCA, FLDA, ICA, HAAR, GF, and an algorithm proposed by Tiwari et al. (2015) that, in fact, fuses the similarity scores from GF and HAAR. Ten samples were collected from 210 infants over two sessions, the first within four hours after birth and another within the next 20-70 hours. Tiwari et al.'s algorithm obtained the highest rank-1 identification rate, 87.78%, whereas PCA, the lowest performing algorithm, had a rank-1 identification rate of 73.27%. Table 2.5 compares the highest rank-1 identification performance for each ear algorithm. To simplify comparison to other infant biometric studies the number of visits and corresponding age are reported as well e.g., 2 (first four hours, 20-70 hours after birth).

2.4.4.1 Performance Summary

Ear recognition for infants appears to perform on-par with ear recognition of adults, with some caveats. The adult database did not has have as many subjects as the infant database and the time between samples is unknown. As show in Table 2.5, it appears that the best performing algorithms for infants tend to be geometric based algorithms: GF, HARR, and the proposed algorithm by Tiwari et al. (2015) which combines GF and HARR. Infants were uncooperative, they often were sleeping or crying. Another obstacle to collecting data was getting willing parents to allow their children to take part in the study.

Table 2.5.
Summary performance results for infant ear recognition

Article	Algorithm	Rank-1 Identification Accuracy	# of Visits (age)
Tiwari et al., 2011	PCA	78.56%	1 (Not Reported)
	FLDA	80.57%	
	ICA	71.75%	
	GF	83.67%	
Tiwari et al., 2012b	PCA	Infants: 81.14% Adults: 83.32%	Infants: 2 (48 hrs apart) Adults: Unknown
	ICA	Infants: 82.42% Adults: 85.13%	
	FLDA	Infants: 87.15% Adults: 89.13%	
	GF	Infants: 89.35% Adults: 91.13%	
	HAAR	Infants: 91.23% Adults: 93.35%	
	LBP	Infants: 90.23% Adults: 92.35%	
	SIFT	Infants: 89.35% Adults: 91.13%	
Tiwari et al., 2015	PCA	73.27%	2 (First 4 hours, 20-70 hrs after birth)
	FLDA	80.62%	2 (First 4 hours, 20-70 hrs after birth)
	ICA	75.32%	
	HAAR	83.24%	
	GF	83.14%	
	Proposed	87.78%	

2.4.5 Fingerprint Recognition

Jain et al. (2014) collected infant fingerprint samples in a controlled lab environment (Michigan State University) at East Lansing, MI and in two health clinics, rural and urban, located in Benin (the city was not mentioned). The lighting at the urban health clinic was fixed in a closed room, whereas the collection at the rural urban clinic occurred at an open-air shelter in sunlight. Fingerprints were collected from infants, 0-4 years old, using the Digital Persona U.are.U 4500, a commercially available device. Both, the left and right, index fingers and thumbs were collected over five sessions, one week apart, at MSU. In Benin, at both health clinics, three left index fingers and thumbs were collected in a single session. Verification and identification performance were evaluated using live-scan and latent fingerprint algorithms from commercially available SDKs. Both verification and identification were evaluated at a baseline, of which, a single template was enrolled per infant. Other matching strategies – to increase performance – were evaluated, namely: the similarity scores from multiple templates were fused; the similarity scores from two fingers were fused together; and templates were updated from additional sessions then the similarity scores were fused from two fingers.

Template updating was only used for identification and the fingerprint samples collected at MSU. The verification rates and rank-1 identification rates from using the latent algorithm outperformed the live-scan algorithm in every matching strategy, regardless of collection location. The latent algorithm performed so well because Jain et al. (2014) observed that infant’s fingerprints and adult latent fingerprints have similar challenges. On average, infants’ fingerprints were of better quality than adults’ according to the NFIQ quality algorithm, although visually the infants’ fingerprints appeared to be of poor quality. The inconsistency between the visual interpretation of quality and image quality algorithm may be because the algorithm was in fact designed for adults. Both fusion methodologies improved the performance rates, and there was noticeable difference in performance between fingerprints collected at MSU and in Benin. The less constrained environment in Benin may have made it more difficult to collect usable fingerprint samples and therefore lowering performance.

Two commercial off the shelf fingerprint SDKs, tenprint and latent, were used to evaluate fingerprint recognition performance using similar matching strategies outlined in Jain et al. (2014), and verification rates and rank-1 identification rates were reported for three age groups: 0-6 months, 6-12 months, and 12 months and older. At an FAR of .1%, the verification rates[4] for the tenprint algorithm improved from 69.19% to 100% for the age groups 0-6 months and 12 months and older, respectively. A similar improvement was seen for the latent algorithm, 73.74% and 100%. The rank-1 identification for infants 0-6 months was 54% and increased to 99% for the 12 months and older age group. However, the latent algorithm had a rank-1 identification of 91% for infants 0-6 months which is higher than what was observed for the latent verification rate and the tenprint identification rate. Moreover, the identification rate for infants 12 months and older improved to 100%. The verification rates and rank-1 identification rates for infants younger than 6 months was significantly lower than those observed for the older ager groups, primarily due to poorer quality samples from younger group. for older infants, primarily due to poorer quality samples collected from the younger group. A custom-made fingerprint sensor designed specifically for infants was created to counter the low-quality fingerprints that were collected with an off-the-shelf commercial fingerprint sensor (Jain et al., 2016). More specifically, the custom fingerprint sensor has a higher resolution of 1270ppi with custom-made dimensions and features to increase the infants' comfort. The custom-made fingerprint sensor was tested by collecting three fingerprint samples from the left and right thumbs over two sessions, 2-4 days apart. The verification rates (at a FAR of 0.1% and 1%) for infants four weeks old and younger were 43.43% and 54.55%, respectively and 79.72% and 83.55% for infants older than four weeks old. The rank-1 identification rates for infants four weeks old and younger and infants older than four weeks were 38.44% and 73.98%, respectively. The verification rates and the rank-1 identification rate for infants older than four weeks were significantly higher than the rates reported for the younger age group.

A longitudinal study by Jain et al. (2017) examined infant fingerprint recognition over time with a 1270ppi custom-made fingerprint sensor described and tested in Jain et al. (2016) and a commercially available device. The study used three sets of data: set A collected three left and right thumb samples using both fingerprint sensors (except the custom sensor in session 1) over a year in four separate session; set B consisted primarily of infants 0-6 months old using only the custom-made sensor in three sessions; and set C also collected primarily from infants 0-6 months using only the custom fingerprint sensor but only over two sessions. Set A collected fingerprint

---

[4] The reported verification rates for infants 6-12 months was interpreted from the reported ROC curves because it was excluded from the text.

samples from 204 infants, 161 completed all four sessions, moreover, subset B had 65 infants and another 40 infants in subset C.

Performance rates were compared for the commercial and custom fingerprint sensors using the 162 subjects in session two and four from set A. At a FAR of 0.1%, the verification rates of the commercial and custom sensor for infants 6-12 months were 95% and 98.9%, respectively and 99.5% and 100% for infants 12-60 months. Infants 0-12 months initially experienced increasing verification rates as more time elapsed between the enrollment and verification images. The verification rates (at a FAR 0.1%) for the commercial sensor at 6 months between enrollment and verification images was lower than the verification rates observed 10 months after enrollment, indicating that image quality could have improved and in turn improving overall performance. The verification rates did not change much between enrollment and verification images with the custom fingerprint sensor, but at a FAR of 0.1%, set B verification rates decreased from 18.0% to 9.8% as the time after enrollment increased from 4 months to 6.

A mixed-effects regression model was used to observe the trend in genuine similarity scores as the time after enrollment and the infant's age at enrollment. The regression model was conducted on the commercially available and custom-made sensor separately. The mixed-effects regression model of the commercially available sensor, with a 12-month time lapse between enrollment and verification images, indicated an increase in genuine similarity scores from 6 to 10 months. The modeled regression line for each age group was parallel which indicated that the age group was not a statistically significant predictor for determining the rate of change in genuine similarity scores over time. Analysis of the mixed regression model for the custom sensor indicated that the mean genuine similarity score was constant between 4- to 6-months after enrollment. The mean genuine similarity scores were significantly different for each age group with significantly lower genuine similarity scores for infants 0-6 months old. The study states that "the mean genuine similarity scores actually increase from 6 to 10 months' time lapse … because the quality of the fingerprints acquired improves as the subject ages" (Jain et al., 2017, p. 1551).

The Dutch Ministry of the Interior and Kingdom Relations (2005) could not capture quality fingerprint samples from infants younger than four. In fact, the capture rate of infants did not surpass 50% until around four years old, and fingerprints could not be captured from infants younger than three. The conclusion of the trial was that obtaining fingerprints from infants was "virtually impossible" (Dutch Ministry of the Interior and Kingdom Relations, 2005, p. 25).

2.4.5.1 Performance Summary

Performance results for infant fingerprint recognition are listed in Table 2.6, additionally, the age, sensor, algorithm, number of visits, and recognition type are listed. After an initial investigation by Jain et al. (2014), two fingers are fused together in subsequent studies because the verifications rates and rank-1 identification rates were higher than the other matching strategies, a single template and fusing multiple templates. Additionally, fingerprints collected outside in an uncontrolled environment performed worse than fingerprints collected in controlled environments. Infants' fingerprints performed poorly on traditional live-scan algorithms compared to a latent fingerprint algorithm, because infants and adults' latent fingerprints exhibit similar challenges. Infants' fingerprints had a better average NFIQ score than adults, creating two discrepancies: 1) visually infants' fingerprints appeared to be of poor quality; and 2) infants' fingerprints performed more poorly than adult fingerprint recognition. To put the performance differences in perspective, an adult fingerprint study, fusing two fingers, had rank-1 identification rates of 90% and 99.8% for the lowest and highest performing algorithms (Watson et al., 2014). Fusing two fingers for infant fingerprint recognition resulted in a range of rank-1 identification rates, 40-90% (Jain et al., 2014).

Capturing infants' fingerprints with adult fingerprint sensors proved to be difficult. Infants' fingerprints are smaller than adults', making it difficult to even acquire an image. The condition of infants' fingerprints is unpredictable and could be dry, wet, or dirty. The custom-made, 1240ppi fingerprint sensor allowed samples with distinguishable ridge spacing to be captured, regardless of age. Infants' fingerprint samples had to be enhanced and similarity scores were fused together yet using an adult fingerprint recognition algorithm was not sufficient (Jain et al., 2016).

Fingerprint recognition is designed for adults. Consequently, the quality assessment algorithms, image enhancement algorithms, matching algorithms, and fingerprint sensors are designed for adult fingerprints. A major assumption of infant fingerprint recognition is that proven fingerprint recognition methods will work for infants. The infant fingerprint literature suggests otherwise: it is difficult to capture fingerprints with distinguishable ridge spacing; infants' fingerprints appear to be of poor quality, but an image quality assessment algorithm indicates infants' fingerprints are better than adults'; and infants' fingerprints require different image enhancement techniques (e.g., adjusting ridge spacing). Therefore, infant fingerprints may in fact possess different salient features for recognition than adult fingerprints.

Table 2.7 describes the results from the longitudinal analysis conducted by Jain et al. (2017), which observed that as time increases between enrollment and verification that the genuine similarity scores of a commercial fingerprint sensor, the Digital Persona U.are.U 4500, temporarily increased and leveled off after an additional two months. More specifically, infants' genuine similarity scores showed an initial increase from 6 to 10 months' time lapse after enrollment. Jain et al. (2017) stated that the initial increase was due to the improvement of image quality. The mixed regression analysis does not include image quality as a predictor suggesting that the claim is based off the median NFIQ2 score which was only reported for three age groups. Or, the claim is purely speculative and based off a visual analysis of subjects' fingerprints. In turn, Jain et al.'s (2017) conclusion is inappropriate and misleading. The performance of the custom-made sensor found that there was no difference in genuine similarity scores two months after enrollment, but there was a significant difference between similarity scores for the three infant age groups. The older the infants were at enrollment the higher the similarity scores.

Table 2.6.

Summary performance results for infant fingerprint recognition

Article	Location	Sensor	Age	Recognition Type	Algorithm	Single Template (baseline)	Fused Templates	Two Fused Fingers	Two Fused Fingers w/ Template Updating	# of Visits
Jain et al., 2014	East Lansing	U.Are.U 4500 (512ppi)	0-4 years	Verification (FAR .1%)	Live-scan	62.25%	71.01%	86,34%	-	5 (1 week apart)
					Latent	78.52%	82.52%	95.04%	-	
				Rank-1 Identification	Live-scan	46.38%	64.16%	73.98%	83.67	
					Latent	75.46%	85.8%	95.52%	98.97%	
	Benin			Verification (FAR .1%)	Live-scan	30.24%	41.67%	57.5%	-	1
					Latent	44.29%	50.24%	64.27%	-	
				Rank-1 Identification	Commercial	20.00%	29.29%	40.00%	-	
					Live-scan	42.86%	55.72%	67.14%	-	
Jain et al., 2015	-	U.Are.U 4500 (512ppi)	0-6 months	Verification (FAR .1%)	Tenprint	-	-	69.19%	-	1
			6-12 months			-	-	94%	-	
			>12 months			-	-	100%	-	
			0-6 months		Latent	-	-	73.74%	-	
			6-12 months			-	-	96%	-	
			>12 months			-	-	100%	-	
			0-6 months	Rank-1 Identification	Tenprint		-	54%	-	
			6-12 months				-	81%	-	
			>12 months			-	-	99%	-	
			0-6 months		Latent	-	-	91%	-	
			6-12 months			-	-	99%	-	
			>12 months			-	-	100%	-	
Jain et al., 2016	-	Custom 1270ppi	<=4 weeks	Verification (FAR .1%)	Not-Specified	-	-	43.43%	-	2(2-4 days apart)
				Verification (FAR 1%)		-	-	54.55%	-	
				Rank-1 Identification		-	-	38.44%	-	
			> 4 weeks	Verification (FAR .1%)		-	-	79.72%	-	
				Verification (FAR 1%)		-	-	83.55%	-	
				Rank-1 Identification		-	-	73.98%	-	
Jain et al., 2017	-	U.Are.U 4500 (512ppi)	0-6 months	Verification (FAR .1%)	Not-Specified	-	-	95%	-	2 (6 months apart)
			12-60 months			-	-	99.5%	-	
		Custom 1270ppi	0-6 months			-	-	98.9%	-	
			12-60 months			-	-	100%	-	

Table 2.7.
Longitudinal performance results for infant fingerprint recognition

Article	Sensor	ppi	Performance Type	Age	Time Lapse (Months)			
					4	*6*	*10*	*12*
Jain et al., 2017	U.Are.U 4500	512	Verification Rate (FAR =.1%)	0-6 months	n/a	66.7%	77.3%	71.1%
				6-12 months	n/a	92.8%	96.2%	94.9%
				12-60 months	n/a	100%	100%	100%
			Rank-1 Identification Rate	0-6 months	n/a	66.7%	77.3%	72.8%
				6-12 months	n/a	99.0%	96.2%	95.8%
				12-60 months	n/a	100%	100%	100%
	Custom	1270	Verification Rate (FAR=.1%)	6-12 months	98.9%	98.9%	n/a	n/a
				12-60 months	100%	100%	n/a	n/a
			Rank-1 Identification Rate	6-12 months	100%	99.4%	n/a	n/a
				12-60 months	100%	100%	n/a	n/a

2.4.6 Iris Recognition

Corby et al. (2006) implemented an iris-based recognition system to identify subjects participating in genetic medical study using the Panasonic Authenticam, a commercial iris recognition device, and Iridian Technologies’s PrivateID V1.5 iris recognition software was used to enroll and identify participants. The study included 1170 subjects, 646 infants and 524 adults. The infants, ranging from 1.5 to 8 years old, were enrolled in the system during their first visit and identified during their second – a year later. The infants’ iris samples were categorized into four groups: full, partial, marginal, or failed enrollment based on both iris samples having acceptable quality, a single iris sample having acceptable quality, both iris samples having marginal quality, and iris samples having unacceptable quality, respectively. Out of the 1170 participants, 184 of them failed to enroll – 155 infants and 29 adults. Furthermore, 495 adults and 491 infants were successfully enrolled at a rate of 94% and 76%, respectively.

Table 2.8 shows the age range, number of infants in each age range (*n*), and their enrollment classification rates (Corby et al., 2006).

Table 2.8.

Infant iris recognition enrollment percentages based on age range (Corby et al., 2006)

Age Range (yr)	*n*	Full Enrollment	Partial Enrollment	Marginal Enrollment	Failed Enrollment
1.5-3	257	26.84%	10.89%	6.23%	56.03%
3-6	295	81.02%	9.83%	5.76%	3.39%
7-8	94	91.5%	6.38%	1.06%	1.06%

The result of an ANOVA test indicated that the infants' mean age for enrollment qualities– acceptable, marginal, and unacceptable were statistically different. Table 2.9 shows the number of infants *n*, the mean age, and age standard deviation σ at a specific enrollment quality.

Table 2.9.

Infant iris recognition enrollment image qualities and mean age (Corby et al., 2006)

n	Avg. Age	σ	Enrollment Quality
457	5.4	0.07	Acceptable
34	4.0	0.25	Marginal
155	2.5	0.12	Unacceptable

The iris recognition system provided the identity of a captured iris which was cross checked with the subject's identification card to ground truth the results; this identification technique is referred to as the rank-1 identification because it returns the identity of the stored template that produced the highest similarity score when matched to a captured iris. A year after the first screening, the iris recognition system correctly identified 488 out of the 491 infants (99.4%) whose enrollment images were classified as acceptable quality. The three infants that were not successfully identified had marginal quality irises – the 31 infants that had marginal enrollment quality images were still identified successfully. The infants that failed to enroll during the first screening half were able to successfully enroll in the second session.

A low-cost iris recognition camera was used as part of a biometric capture system to enroll and verify refugees across multiple locations – adults and infants were both enrolled (Bachenheimer, 2016). The device is currently $480.00 (Fulcrum Biometrics LLC, 2018). The iris recognition device failed to capture irises 85.9% of the time from infants younger than four years old and 2% for infants older than four. Moreover, the majority of infants, approximately 1%, had two low-quality irises. Bachenheimer (2016) mentioned that the iris device could have had a high failure to capture rate for the younger age group, because the device, which resembles binoculars, had to be directly held up to the infant's face long enough to capture an iris. Additionally, he states that a more usable iris device may improve the capture rate.

2.4.6.1 Performance Summary

The literature for infant iris recognition is limited. However, studies do indicate that iris images of good quality can be captured – only Bachenheimer (2016) and Corby et al. (2006) have conducted iris recognition research with infants. Furthermore, iris recognition can correctly identify 99.4% of infants a year later, that is, of the samples that could be captured. Both studies' results show that capture rates for young infants are fairly low and improve substantially with age. When examining infant iris recognition, the device and age could significantly impact performance. It is important observe these factors when examining infant iris recognition. It is interesting to note the infant iris recognition literature does not mention any evidence to support that the iris pattern does not stabilize for the first two years after birth – which other authors have unsubstantially claimed (Barra et al., 2014, 2014; Jain et al., 2004; Jia et al., 2012; Tiwari et al., 2015, 2013, 2013; Tiwari & Singh, 2012; Weingaertner et al., 2008).

2.4.7 Multimodal Biometrics

Tiwari et al. (2012) observed that a rank-1 identification rate for face recognition was 80.42%. By combining the infant's face with additionally collected soft biometric data, the rank-1 identification rate improved. The rank-1 identification rates were evaluated with the face combined with all the collected soft biometric data and a single soft biometric. Adding an infant's sex, blood group, weight, and height increased the rank-1 identification rate by approximately 2%, 3%, 2%, and 3%, respectively. When the face was combined with all four soft biometrics the identification rate increased by 6%.

Infant ear recognition had a rank-1 identification rate of 83.67% and when fused with a single soft biometric (sex, blood group, weight, and height) improved by approximately 2%, 2%, 3%, and 3%. When the ear was fused with all four soft biometrics, the identification rate improved by approximately 6% (Tiwari et al., 2012c). The improvements to rank-1 identification rate improved by the same rate when fusing face with all four soft biometrics. However, ear recognition, by a small margin, outperformed face recognition.

Madhu et al. (2017) combined the footprint of an infant and its mother's fingerprint to improve performance. The similarity scores from the infant's foot and mother's fingerprint were fused independently. The fusion methodology achieved an FNMR of 12.3% at an FMR of 0.01%.

2.4.7.1 Performance Summary

Multimodal biometrics could be useful for improving recognition accuracy for infants when little information can be extracted to discriminate between infants' biometric samples. Combining an infant's biometric with their mothers can also increase recognition accuracy. Although soft biometric data is easy to record, an infant's height and weight can change. Additionally, multimodal techniques combining the infant's and mother's biometrics depends on the mother always being present to successfully identify or verify an infant's identity. The mother may not always be present, therefore combining an infant's biometric with its mother is not realistically sustainable. Table 2.10 summarizes the studies that use multiple biometric modalities for infant recognition.

Table 2.10.
Summary of Multimodal Biometrics

Article	Performance Type	Fused Biometrics	Performance
	Rank-1 Identification Rate	Face + Sex	82%
	Rank-1 Identification Rate	Face + Blood Group	83%
Tiwari et al., 2012	Rank-1 Identification Rate	Face + Weight	82%
	Rank-1 Identification Rate	Face + Height	83%
	Rank-1 Identification Rate	Face + All Four	86%
	Rank-1 Identification Rate	Ear + Sex	85.12%
	Rank-1 Identification Rate	Ear + Blood Group	85.16%
Tiwari et al., 2012c	Rank-1 Identification Rate	Ear + Weight	86.16%
	Rank-1 Identification Rate	Ear + Height	86.46%
	Rank-1 Identification Rate	Ear + All Four	89.26%
Madhu et al., 2017	Identification (0.01% FMR)	Mother's Finger + Infant's Foot	87.7%

2.5 Challenges of Infant Biometrics

Infant biometrics, regardless of modality, exhibit special or exaggerated challenges due to the natural non-cooperative behavior of infants (Bharadwaj et al., 2010; Corby et al., 2006; Jain et al., 2015, 2016, 2017, 2014; Jia et al., 2012; Lemes et al., 2011; Tiwari et al., 2015; Weingaertner et al., 2008). Other challenges suggested by Jain et al. (2014), Jia et al. (2010), Joun et al. (2003), and Lemes et al. (2011) arise due to inherit traits of the biometric at infancy, and the unique challenge of using a device that is designed for and used by the adult population (Bachenheimer, 2016; Jain et al., 2014; Weingaertner et al., 2008).

Infants' fingers are known to be excessively oily or wet from natural characteristics or from behavior e.g., placing fingers in their mouth. In addition infants tend to keep their fists balled and may become agitated when they are opened (Jain et al., 2014). The ridge spacing of infants' fingerprints is smaller than an adult which must be adjusted to match the ridge spacing of an adult's fingerprint before extracting features (Jain et al., 2015, 2014; Joun et al., 2003). A fingerprint sensor may not detect an infant's finger due to its small size compared to an adult for which the sensor was designed. A high resolution custom made fingerprint sensor has shown promise by mitigating challenges due to ridge spacing and wet or oily fingers (Jain et al., 2017). Adults are primarily the target population of fingerprint recognition. Quality assessment algorithms, image enhancement algorithms, matching algorithms, and fingerprint sensors are designed for adult fingerprints. Infant fingerprint literature posed unique challenges: it is difficult to capture fingerprints with distinguishable ridge spacing; infants' fingerprints appear to be of poor quality, but an image quality assessment algorithm indicates infants' fingerprints are better than adults'; and infants' fingerprints require different image enhancement techniques (e.g., adjusting ridge spacing). Adult fingerprints may in fact possess different salient features than infant fingerprints, essentially rendering adult fingerprint matching algorithms, image quality assessment algorithms, and fingerprint sensors practically unusable for infant fingerprint recognition.

Palmprint recognition and footprint recognition both use ridge based biometric which is also used in fingerprint recognition. Therefore, they exhibit the same challenges as fingerprint recognition. Due to the characteristics of infant's skin it was difficult to apply the right amount of pressure to mitigate deformation of the palm's ridges. Infants would also get extremely irritated due to hunger or tiredness and would often cry making it difficult to capture palm images (Jia et al., 2012). Footprint recognition also had difficult capturing usable ridges and had to test multiple sensors which all resulted in poor quality samples (Weingaertner et al., 2008). Ridge based biometric matching and quality assessment algorithms are designed for adults, suggesting that important features for palmprint and footprint recognition may be different for adults and infants.

Face recognition, which is sensitive to facial expression, typically requires images with neutral expressions. It was challenging to obtain face images with neutral expressions from infants. Infants were consistently crying or sleeping and had difficulty fully opening their eyes, making it difficult to detect their face. Typically, the eyes are used in face detection algorithm. In fact, that would indicate that infants' and adults' faces possess different distinguishing features necessary for successful face recognition. Infant's also had difficulty keeping still which could introduce motion blur into the image (Bharadwaj et al., 2010). Ear recognition was also challenged by infant movement making it difficult to capture good quality ear images (Tiwari et al., 2012a).

Iris recognition exhibited similar problems to other biometrics. According to Bachenheimer (2016), it was difficult to capture images because of the usability of the device which had to be held up to an infant's face to capture iris images. Infants also exhibited difficulty properly positioning their head, keeping it still, or opening their eyes making it difficult to capture iris images (Corby et al., 2006). About half of the infants during enrollment were younger than four years old and consequently had the lowest image qualities and capture rates. Regardless, 99% of the infants that could be enrolled were correctly identified one year later. Table 2.11 indicates that most challenges associated with infant biometrics stem from their uncooperative behavior. Each biometric has difficulty getting infants to cooperate, leading to issues of capturing a biometric. All biometrics are susceptible to sleeping, crying, and screaming all of which making correctly positioning a biometric more difficult and can also lead to more subject movement. Ridge based biometrics had difficulty overcoming wet and oily fingers. The size of an infant's fingerprint and ridge spacing also makes it difficult to capture infant fingerprint samples using devices made for adults. This is because the sensors are designed, ergonomically and algorithmically, for adults' fingerprints, and the sensors are expecting larger fingerprints and ridge spacing. Creating a custom-made fingerprint sensor for infants seemed to mitigate some of these challenges. Iris recognition devices are made for adults, however, the challenges associated with the device do not seem to be physical, like fingerprint recognition, but seem to be associated with the usability of the device. Bachenheimer's (2016) device required the infant's head to be pressed against the device, whereas, Corby et al.'s (2006) device required infants to be some distance away, approximately 19-21 inches. However, there have not been enough infant iris recognition studies to conclude anything about the device but there is some anecdotal evidence to support a more thorough investigation.

Table 2.11.

Infant Biometric Challenges

Challenge	Face	Finger	Foot	Iris	Palm	Ear
Oily and Wet Skin	-	X	-	-	X	-
Balled Fists	-	X	-	-	X	-
Ridge Spacing	-	X	X	-	X	-
Facial Expression	X	-	-	-	-	-
Difficulty Keeping Still	X	X	X	X	X	X
Properly Positioning Biometric Characteristic	X	X	X	X	X	X
Acquisition Device Was Designed for Adults	-	X	-	X	X	-
Crying	X	X	X	X	X	X
Sleeping	X	X	X	X	X	X
Screaming	X	X	X	X	X	X

2.6 The Eye

This section outlines the basic structure and formation of the eye. Furthermore, it also gives a detailed account of the iris structure and describes the features in the iris that contribute to its uniqueness. To understand the unique challenges of infant iris recognition, this section also gives a detailed account into the development of the iris.

2.6.1 Structure and Formation of the Eye

The eye, displayed in Figure 2.3, has three layers: the outer layer, uvea (i.e. the middle layer), and the inner neural layer. The outer layer's two main components are the cornea and the sclera which are made of collagen fibers that assist in protecting the inner parts of the eye. The primary function of the cornea – which is transparent – is to refract light onto the retina while the sclera – the white opaque area of the eye – is the dense, white, fibrous tissue that surrounds the iris (Bridges, 2015; Oyster, 1999; Remington, 2005).

The middle layer of the eye, listed from the posterior to the anterior, is comprised of the choroid, ciliary body, and iris. The choroid is made up of blood cells and melanin pigments that absorb light to prevent the scattering of light inside of the eye. The ciliary body is next to the lens and includes the ciliary muscle that controls the shape of the lens. Furthermore, the ciliary body assists in producing parts of the aqueous humor. The iris is the colored area that is visible through the cornea. There are two muscles in the iris that help control the size of the pupil, the sphincter and dilator muscles (Bridges, 2015; Oyster, 1999; Remington, 2005).

The inner layer of the eye contains three parts: the anterior chamber, posterior chamber, and vitreous chamber. The anterior and posterior chamber are connected through the pupil which contains aqueous humor. The vitreous chamber contains the vitreous humor which is a gel-like substance. Additionally, the inner layer of the eye contains the retina which detects light and sends information to the brain through the optic nerve (Bridges, 2015; Oyster, 1999; Remington, 2005).

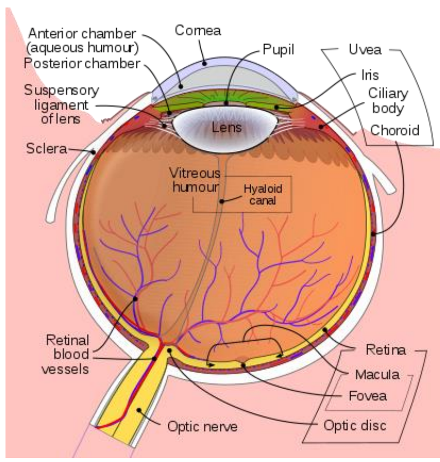

Figure 2.3. Structure of the Eye (Rhcastilhos, 2018)

In the third week of gestation, the primary germ layers are formed, and the development of eye structures begin with the ectoderm and mesoderm. A month into the embryonic period, the eye begins to develop and within another month it develops into a miniature version of the adult eye, with some basic elements. The eye begins to develop its important structures such as the cornea, lens, optic nerve, and retina six weeks into gestation and the eye is roughly two thirds of its final size. The optic cup and optic stalk are the beginning of the retina and optic nerve. The outer rim of the optic cup develops into the epithelial layers for the iris, ciliary body, and iris muscles. The iris is complete five months into the gestation period but the epithelial layers do not progress to the center which causes the pupil to not be fully formed until seven months into the gestation period (Oyster, 1999; Remington, 2005).

The cornea begins to form after the first month of gestation. All the layers of the corneal epithelium are complete by the sixth month and all its structures are complete by the end of the seventh month. The cornea is almost fully grown at birth and will finish growing within the first couple years. Although the structure is fully formed, an infant's cornea is thicker and more curved than an adult's. The cornea accounts for about 15% of the surface of an adult's eye and 25% for an infant's (Oyster, 1999; Remington, 2005).

The iris begins to form around the third month of gestation, which begins as the outer layer of the optic cup (Remington, 2005). The iris sphincter muscle begins to form in the fifth month and both, the dilator and sphincter muscles, is fully developed before birth (Remington, 2005). Pigmentation in the anterior epithelium and dilator muscle begins to appear at week 10 and are complete by the end of the seventh month (Remington, 2005). The formation of the anterior border layer and stroma are completed before birth; according to Oyster (1999), "viewed from the front, the iris is nearly complete by the end of the fifth month of gestation, with recognizable muscle and epithelial layers, blood vessels, and so on, but it still lacks a pupil" (p. 442). The epithelial layers have not completely converged to the center of the iris and will do so at 7 months into gestation. The pigmentation of the stroma and anterior border layer continues to develop after birth and varies significantly, and it is the most significant change of the iris after birth (Oyster, 1999; Remington, 2005).

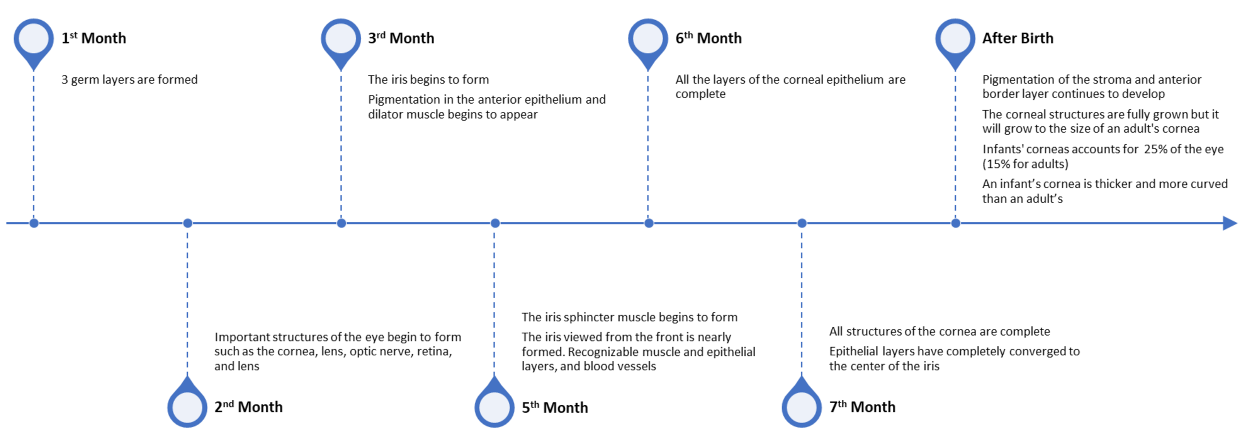

Figure 2.4. Development of the human eye timeline

2.6.2 Structure and Surface of the Iris

The iris is divided into four layers. From the posterior portion of the iris to the anterior the layers are: the posterior epithelium, the anterior epithelium and dilator muscle, the stroma and sphincter muscles, and the anterior border layer – sometimes the border layer is grouped with the stroma. The posterior epithelium is a single layer of pigmented cells which curls around to the surface of the iris, encircling the pupil, which forms the pupillary ruff. The anterior iris epithelium is anterior to the posterior epithelium and lies closest to the stroma of the iris. The top portion of the anterior iris epithelium is pigmented, and the bottom portion is made up of muscle processes. The dilator muscle – runs from the midportion of the sphincter muscle to the iris root – consists of radial fibers, when dilated they pull the pupillary portion of the iris outwards in the direction of the iris root (Remington, 2005).

The iris stroma is made up of connective tissue which contains collagen fibers and cells that are pigmented and non-pigmented. Within the stroma lies the sphincter muscle which is a circular muscle in the pupillary zone, and it constricts the pupil when the muscle is contracted. The anterior border layer is composed of interweaving meshwork with fibroblasts directly on the surface and pigmented melanocytes below. The melanocyte layer's characteristics vary among irises and contribute to iris color – the meshwork density, arrangement, and thickness. Additionally, the collagen fibers are arranged radially and weave between the melanocytes and fibroblasts which can also be seen in lighter colored irises. Iris crypts are the areas of the iris which do not have the anterior border layer (Remington, 2005).

The color of the iris comes from the density of the anterior border layer and stroma's tissue, pigment density in a melanocyte, and the density of the melanocyte itself. If an iris is light the collagen fibers are visible, whereas dark irises appear smooth from the density of the anterior border layer (Remington, 2005).

The iris surface has several distinct features such as the crypts (i.e. crypts of Fuchs), collarette, radial furrows, and concentric furrows (U.S. Patent No. 4,641,349, 1987). The iris is divided into two areas the ciliary area and pupillary area which divides the iris from the collarette to the pupil and the collarette to the outer boundary of the iris, respectively. The basic structure of the iris contains the posterior and anterior layers. The posterior layer of the iris is darkly pigmented, and the anterior layer's pigment ranges from light to dark but never reaching the same level of darkness as the posterior layer of the iris. The anterior layer of the iris has strands of tissue that weave and create gaps and holes which are referred to as the crypts of Fuchs. The crypts of Fuchs vary for all irises, and contribute to the individuality of the iris, they can be used as a unique identifier (U.S. Patent No. 4,641,349, 1987; Oyster, 1999). The collarette is the area that lies between the ciliary and pupillary areas and has a wave shaped line. The radial furrows are creases in the tissue and bulge outward which allows the iris to dilate or contract to control the size of the pupil. The creases in the tissue extend out like rays of light from the pupil through the collarette. The concentric furrows appear close to the outer boundary of the iris and are creases in the tissue that bulge outward in a circular manner. The concentric furrows assist in the expansion and contraction of the iris in different directions than the radial furrows (U.S. Patent No. 4,641,349, 1987). Figure 2.5 displays a diagram of the regions and distinct features that make up the iris.

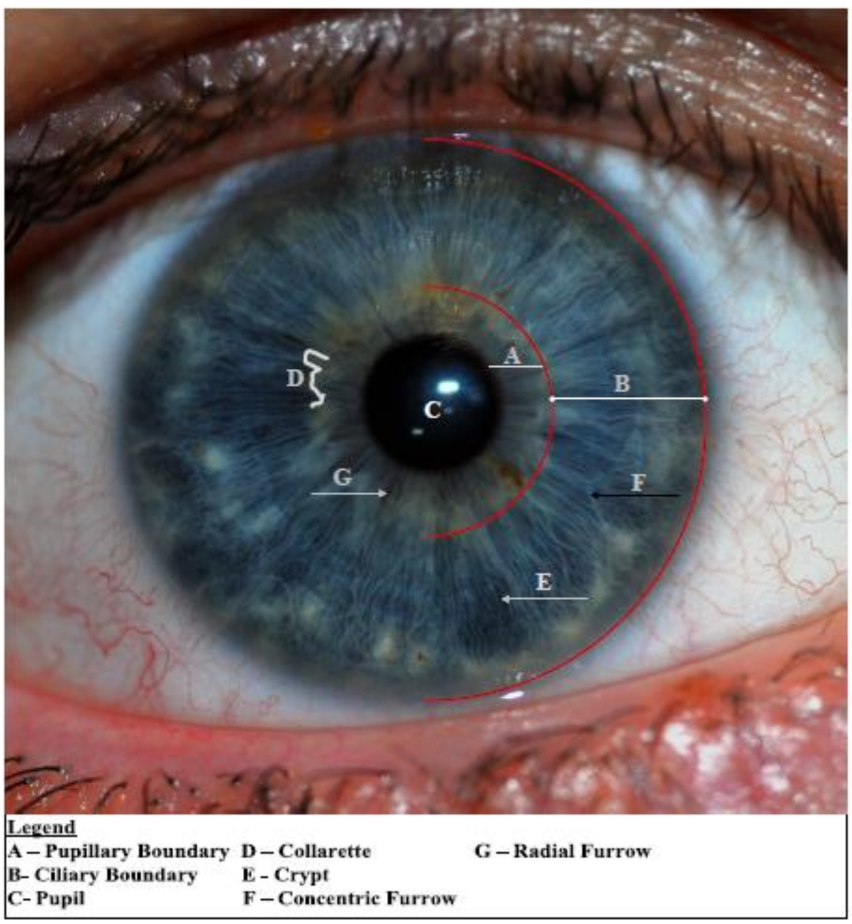

Figure 2.5. Annotated iris image displaying characteristics and features of the iris. This image was modified from the original image (Drewes, 2007)

2.7 Iris Recognition

Iris recognition is a method of biometric authentication that uses the pattern of the iris to identify an individual. The iris is the colored area of the eye, which is externally visible, but is an internal organ which is well protected from damage. Resistance to damage makes the iris an ideal biometric compared to biometrics that are more susceptible to damage such as fingerprints. Some believe that iris patterns are stable over time, even from birth.

To perform iris recognition, the iris must first be segmented from the rest of the eye. Segmentation is done by detecting the boundary of the pupil and the boundary that separates the sclera and the iris. Next, the iris must be interpolated by remapping each point in the iris region from a Cartesian coordinate system to a polar one. This remapping automatically normalizes the area from the pupil boundary and iris sclera boundary of the iris. Normalizing the iris reconciles any deformation of the iris due to constriction or dilation of the pupil and makes iris recognition mostly resistant to changes in size of the iris. The iris code is generated from the normalized iris image by extracting phase information of the iris pattern. A masking code is calculated to indicate the area of where iris obstructions are located and circumvents errors from obstructing features such as eyelids, eyelashes, and specular reflections. To compare two irises to each other, a similarity or dissimilarity score between the two irises is calculated by using the two iris codes and masking codes generated during feature extraction (Daugman, 2004).

2.7.1 History of Iris Recognition

One of the earliest recorded accounts for using irises to recognize individuals was in 1886 and was implemented to identify repeat offenders in France (Bertillon, 1886). Flom and Safir obtained a patent for the first iris recognition framework which described the use of an iris's unique features for the identification of individuals by comparing an obtained image to stored reference images (U.S. Patent No. 4,641,349, 1987). After Flom and Safir's patent was published, a study was conducted to determine if the use of an iris to recognize individuals is feasible and if the features of the iris remain stable. The study collected approximately 1000 iris images from 650 individuals and concluded that the iris pattern is stable over time, moreover, the iris pattern is unique between an individual's left and right eyes and between individuals (Johnston, 1991). Daugman was granted a patent in 1994 for developing the first operational iris recognition system.

Daugman's approach to iris recognition is an influential step in iris-based biometrics and remains a primary driver in iris biometric technology today. The company that owned the patent rights to Flom and Safir's iris recognition framework also owned Daugman's patent for the first operational iris recognition system (U.S. Patent No. 5,291,560, 1994; U.S. Patent No. 4,641,349, 1987).

2.7.2 Iris Acquisition

All commercial iris recognition systems follow these basic principles: illumination from controlled and ambient light sources, a camera and light source from a standoff distance, acquisition of the iris image through the camera, and then the iris is segmented, normalized, and generated into an iris code – proprietary software such as Neurotechnology may use other methods apart from iris code. Iris systems can capture one or both irises at the same time. A good quality iris should have a resolution of at least 60 pixels or more across which may require some devices to be in very close proximity to subjects (JTC 1/SC 37, 2011). Typical commercial iris recognition devices require cooperation from its users; slight movements from subjects could produce motion blurred iris images. Iris on the move attempts to offset reduced subject cooperation while continuing to capture high-quality iris images (Matey et al., 2006).

The reflectivity of the iris is dependent on the wavelength of ambient and controlled light sources. Near-infrared reflectance (NIR) illuminators produce the best reflectivity of the iris which in turn reveals rich iris features, even for darkly pigmented irises. A wavelength of 700-900nm is typically used in most iris recognition systems (Ackerman, 2016). There are three types of iris recognition devices: NIR cameras, high-resolution color cameras, and telescope cameras. The NIR camera illuminates the iris at the wavelength 700-900nm and typically captures at short distances which requires cooperation from users of the iris recognition system. NIR cameras are most commonly used because of its ability to distinguish features and textures for darkly pigmented irises. High-resolution color iris cameras are typically used in research to analyze iris patterns and require very high-level cooperation from users because acquisition takes place at a very close range. Getting well defined features from the iris in the color spectrum is difficult especially for darkly pigmented irises because light is absorbed and not reflected as well as lighter colored irises. Telescope iris cameras can capture irises at long distances (i.e. 10ft) and use a stronger NIR illuminator than typical NIR cameras. Telescope iris cameras also have strong de-blurring capabilities to enhance images that make it ideal for capturing irises from non-cooperative users (Du, 2006).

2.7.3 Iris Segmentation

Iris acquisition captures an image of an eye which includes features such the sclera, pupil, eyelids, and eyelashes. Segmentation locates and removes just the iris from the rest of the eye by detecting the boundaries of the pupil and iris and removing occluded portions of the iris (Jillela & Ross, 2016; Roy & Soni, 2016).

Several factors can impact the accuracy of iris segmentation such as occlusion and illumination. Inaccurate segmentation of the iris can in turn degrade performance of an iris recognition system. Eyelids or eyelashes that occlude the iris can make segmentation difficult, but most segmentation processes aim to additionally detect the boundaries of eyelids and eyelash occlusions and remove them during segmentation. Illumination can also be problematic for iris segmentation. Low contrast between the boundaries and iris region can be caused by poorly illuminated irises, and specular reflections can occur from poorly aligned illuminators. Poor illumination makes it difficult to distinguish iris textures from each other and specular, reflections near the boundaries cause high intensity pixels in the iris resulting in abrupt changes in pixel values. Iris segmentation accuracy is also affected by lack of user cooperation which can cause off-angled iris images and motion blur. Motion blur can also be caused by moving cameras or the eye itself. The most common iris segmentation algorithms are the integro-differential operator (Daugman's classic approach) and the Hough transform (Jillela & Ross, 2016).

2.7.4 Iris Normalization

The iris changes and becomes deformed when the pupil constricts and dilates. This causes iris features to become unaligned if images are acquired in different conditions that could cause variation in pupil size. Iris normalization attempts to compensate for this deformation and remaps the segmented iris to account for variations in scale and rotation of iris features. The most popular method is Daugman's rubber sheet model that transforms the segmented iris from a Cartesian polar coordinate system to a dimensionless polar coordinate system and normalizes the scale of the iris. Rotation variations are accounted for during matching by selecting the best matching results from shifting the rubber sheet model's x-axis which represents the rotation of an iris in Cartesian coordinates. Other methods of iris normalization account for iris deformation by using image registration techniques. A newer method of iris normalization uses non-linear transformations to better model an iris's natural deformation response from different lighting intensities and the distribution of iris muscles that control the constriction and dilation of the pupil (Thainimit, Alexandre, & de Almeida, 2013).

A proposed non-linear iris normalization technique combines a non-linear transformation and linear unwrapping of the iris to normalize the iris images. Using the ratio of the inner and outer boundaries of the iris, a reference ratio is calculated for all iris images to be scaled to. Connecting a point on the pupil boundary with another on the outer iris boundary creates an arc that changes in length (angular direction) and radius. However, the changes in angular direction are ignored and the radial changes are favored to deform the image nonlinearly to the referenced annular zone. After the iris is transformed nonlinearly it is unwrapped to linearly fit a fix-sized rectangular model (Yuan & Shi, 2005).

2.7.5 Iris Feature Extraction

An iris code is a mathematical representation of the extracted iris features. Creating an iris code according to Daugman's method is done by demodulating an iris image using 2-D Gabor wavelets. The quadrant that the phasor of the 2-D Gabor wavelets lie in determines a 1 or 0 depending on the sign of the value returned by the wavelet function. The iris code is cyclic meaning that it is represented by a single bit change. The bit stream is 2,048 bits long. Masking bits are also computed to signify areas of the iris that are occluded from eyelids, eyelashes, or poor signal noise ratio. The independence of two iris codes is calculated as a Boolean logic exclusive-or, in which, the hamming distance is calculated and represents the similarity between any two irises. The lower the hamming distance between two irises the more similar they are to each other (J. Daugman, 2004).

2.7.6 Summary

The process of iris recognition includes acquiring an image of the eye, segmenting the iris, normalizing the iris image to account of scale and rotation variations, and create an iris code that is used to match two irises together. Iris acquisition can be problematic for non-cooperative users because in general iris devices require users to be near to and looking at the camera, and remaining still. Some of these issues can be addressed with telescopic iris devices that acquire iris images quickly and retrieve iris images of the highest quality. Having too little or too much illumination can make it difficult to distinguish important features of the iris and cause segmentation or feature extraction failure. Uncooperative subjects can also impact segmentation accuracy by not looking directly at the iris camera which causes off-angled iris images making proper segmentation difficult. Furthermore, uncooperative subjects can also cause blurred images by moving during acquisitions or severe occlusions resulting from eyes that are not fully open. Blurry images can also be caused by movement of the iris camera or the eye itself.

2.8 Iris Aging Effects

Decreased genuine match scores overtime is a phenomenon known as template aging (JTC 1/SC 37, 2017). In iris recognition an template aging affect occurs when “the quality of the match between an enrolled biometric sample and a sample to be verified degrades with increased elapsed time between samples” (Fenker & Bowyer, 2011, p. 232). Whereas, an iris aging effect “would be some definite change in the iris texture pattern due to human aging” (Fenker & Bowyer, 2011, p. 232). Iris template aging research is dedicated to determining what factors cause genuine similarity scores to change over time. In the iris recognition literature, aging effects are a contentious topic. Some literature claims that they observed an aging effect and continues to do so. However, those claims have been heavily disputed (Grother, Matey, & Quinn, 2015; Grother et al., 2013; Mehrotra, Vatsa, Singh, & Majhi, 2013; Sazonova et al., 2012; Trokielewicz, 2015), primarily because of large variations in the pupil-to-iris ratios over multiple samples which, in turn, lowered performance.

Another definition for iris ageing is “irreversible changes to the healthy iris or neighboring anatomy that yield mated dissimilarity scores that increase monotonically with time-separation of compared images” (Grother et al., 2013, p. 9). This definition of iris aging is dependent on the use of a biometric matching algorithm to detect permanent changes in the iris. Different from the definition of template aging, iris aging requires a permanent change in the iris or neighboring anatomy and that the similarity scores would continuously decrease. For example, large variations in pupil-to-iris ratios over multiple samples would be caused by variation in lighting. Therefore, the changes in genuine match scores would not be permanent and would not be considered as an iris aging effect. Another study emphasized that an aging effect would cause genuine similarity scores to continuously and gradually decrease (Mehrotra et al., 2013).

In addition to iris aging and template aging, iris stability is another metric that examines changes in an iris's genuine and impostor similarity scores over time (Petry, 2015). The Stability Score Index was developed by (O'Connor, 2013), and is used to understand how much an individual's genuine similarity scores and impostor scores vary over two different samples in reference to the maximum variation for all individuals. Examining the stability of adults' irises over 6 visits and one month apart resulted in a slight change in similarity scores but the stability score index did not change. In conclusion, with samples collected one month apart the iris remained statistically stable over all 6 visits (Petry, 2015).

Biometric permanence is another metric that can be used to determine the stability of genuine similarity scores over time. The metric considers causes of variability by examining the difference between intra-visit and inter-visit genuine similarity scores. Therefore, the change in genuine similarity scores would then be due to only template aging. Biometric permanence is measured as a ratio of the complement FNMR after some time-frame and the complement FNMR at verification (Harvey, Campbell, Elliott, Brockly, & Adler, 2017). Iris or template aging affects has some challenges which biometric permanence attempts to solve (Fenker & Bowyer, 2011; Grother et al., 2015, 2013).

Healthy individuals' genuine match scores may vary because of changes in the sensor, environment, subject behavior, or the physical iris itself (Grother et al., 2013). Iris camera optics can degrade over time, potentially increasing error rates (Bergmuller, Debiasi, Uhl, & Sun, 2014). Changes in lighting condition or an iris camera's illuminator can directly affect genuine match scores. Environmental related effects can be due to changes in ambient or infrared illumination to a user's iris. The lack of cooperation of a subject or increased familiarity with a device can also impact genuine match scores. The iris itself can also exhibit changes, permanent or temporary, that could cause genuine match scores to change. The temporary changes of genuine match scores can be attributed to changes in environmental conditions, device characteristics, or subject behavior. Permanent changes would be reflected by irreversible changes to the iris and surrounding anatomy of the eye. For a healthy individual, physical eye changes may be seen in the cornea shape, iris texture, or natural pupil dilation changes that can occur over a person's lifetime (Grother et al., 2013).

A longitudinal study examining iris aging for adults in an operational scenario used a general linear mixed model because it can handle "multiple responses that are imaged irregularly over time, and potentially correlated over time…fixed effects model population-wide variation… random effects give subject-specific regression effects" (Grother et al., 2013, p. 26). Dissimilarity scores increased over time but, the rate of change was a magnitude less than what is expected for an average individual's expected variation. Some individuals experience a greater increase in dissimilarity scores.

The same study, using datasets collected by researchers at Notre Dame, concluded that the observed performance degrades with increased time between matched samples because of varying environmental conditions which caused variations in pupil dilation, usable iris area, and their joint effect. The generalized linear mixed effects regression modeled the dilation and usable iris area effects on genuine match scores to obtain an individual specific rate of change. The modeled effects were subtracted from the observed genuine match score for each pairwise match and the false non-match rate performance was revaluated. The adjusted performance for dilation and usable iris area affects exposed an absence of a detectable age effect (Grother et al., 2013).

An additional study, using robust regression, determined that local contrast, occlusion, illumination, and sharpness were all significant predictors in the regression model. Given the four quality factors, the elapsed time between samples was still a significant predictor for the regression model. Therefore, the elapsed time between samples are significant for observed changes in genuine match scores and, in part, attributed to image quality metrics. The researchers also mention that the change of genuine match scores over time could be caused by pupil dilation or sensor aging (Sazonova et al., 2012). A multiple linear regression analysis also concluded that the time elapsed between matched sample is significant and image quality metrics are also significant factors for modeling the variation of match scores. The inclusion of image quality metrics in the model lowered the rate of change due to elapsed time and significant predictors varied depending on the iris recognition algorithm used. Aging effects may be separate from image quality but image quality metrics should remain in models because of their observed significance for impacting genuine match scores (Trokielewicz, 2015).

Another study observed an aging affect that caused a degradation of high genuine match scores from elapsed time between samples. A correlation between mean pupil-to-iris ratios and genuine match scores was not observed. Additionally, no physical changes to the iris texture were observed (Baker, Bowyer, & Flynn, 2009). Averaging the differences in pupil-to-iris ratios may hide high variations between pupil-to-iris ratios thus, explaining why no correlation was observed (Grother et al., 2013). It is has been debated by the research community in more than one occasion that the observed aging effect in the Notre Dame studies are due to pupil dilation differences between matched pairs (Grother et al., 2013; Mehrotra et al., 2013).

Another study attempted to control for changes in pupil dilation by excluding images that had an observed dilation greater than 0.1. The increase of dissimilarity scores differed between algorithms, concluding that there is a template aging effect but that it was smaller than what was observed in other studies because large changes in pupil dilation were excluded from analysis (Fenker & Bowyer, 2011). When examining intra- and inter-session error rates from four sessions ranging from one to four weeks apart, the false reject rate increased as time between samples increase, leveling off in the fourth and final session (Tome-Gonzalez, Alonso-Fernandez, & Ortega-Garcia, 2008).

CHAPTER 3. METHODOLOGY

This study determined if iris recognition performance for infants between the age 0-2 years old is feasible by answering three main research questions: 1) is there a difference between image quality metric scores for adults and infants; 2) is there a difference in matching performance for different age groups; and 3) do genuine similarity scores change over elapsed time? This study analyzed adults because performance and image quality results are well known for this population. Therefore, adults were used as a baseline when examining the performance and image quality of iris recognition for infants.

3.1 Infant Data Collection

The data used in this study were captured in multiple visits as part of a longitudinal multimodal collection on infants. Thus, this data is used in secondary analysis.

3.2 Adult Data Collection Methodology

Again, to compare with infants, the data used in the secondary analysis came data came from an existing dataset collected in 2013, for more details see (Petry, 2015). Only one of the 6 visits from the adult data collection was used in this study.

3.3 Iris Camera

Both the adult and infant irises were collected with the same iris camera. The camera is stationary and can capture irises from up to 8 feet away. It also captures the left and right iris and face of an individual simultaneously; iris images are captured at the NIR spectrum (AOptix, 2011). Table 3.1 summarizes additional parameters of the used in this study.

Table 3.1.
Iris Camera Specifications

Parameter	Value /Functionality
Stand-off distance range	4.9-8.2ft
Image capture cycle time (2 irises and face)	4 seconds
Illumination	820-860nm (NIR)
Capture volume	1ft deep, 3.3ft high x 2.46 wide at a standoff of 6.6ft
Dual-iris capture	Yes

3.4 Hypothesis 1

There is no difference between image quality metrics for adults and infants. To address this hypothesis, a comparison between means of four groups were conducted for each image quality metric. The four groups were infants 0 to 6 months old, 7 to 12 months old, 13 to 24 months old, and adults. These infant age groups were selected because they were similar to what was chosen in a longitudinal infant fingerprint recognition study by Jain et al. (2017); the age groups used by Jain et al. were 0-6 months, 6-12 months, and 12 months and older. The Neurotechnology 10 SDK was used to compute the image quality metrics and extract templates from the raw images. A quality assessment algorithm may fail to compute quality metrics, these images were removed from analysis.

3.5 Hypothesis 2

There is no difference in matching performance for different age groups. Four groups were compared: infants 0 to 6 months, 7 to 12 months, 13 to 24 months, and adults.

3.6 Hypothesis 3

This hypothesis determined if the time between samples had a significant effect on genuine similarity scores. To determine if elapsed time is a significant predictor a linear mixed model was used. Other image quality metrics were used in the model to obtain an adequate fitting model. Also, the additional image quality metrics gave a comparison for the effect that elapsed time had on genuine similarity scores in relation to the other metrics in the model. If the genuine similarity scores did change over elapsed time, then the null hypothesis ($\beta_{\Delta T} = 0$) was rejected in favor of the alternative hypothesis ($\beta_{\Delta T} \neq 0$).

3.7 Generalized Linear Mixed Model

A linear regression model consists of coefficients that are considered fixed and explain a population-wide variation. However, in some cases it may be necessary to incorporate random effects, especially if the observations are correlated. A regression model that has both random and fixed effects is considered a mixed model (Jiang, 2007).

CHAPTER 4. RESULTS

The analysis is divided into three sections: analysis of image quality metrics and performance between age groups; and an analysis of genuine similarity scores and if they change over time.

4.1 Data Cleaning Procedure

Image quality metrics were processed using Neurotechnology 10 SDK. Irises that failed to compute quality were removed. Subsequently, templates were created. The settings of the Neurotechnology 10 SDK were set so that all images produced templates. Four samples were removed due to a processing error, in which the resolution of the images was abnormal resulting in very small iris images. After cleaning there were 233 images for the 0-6 months group, 479 images for the 7-12 months group, 541 images for the 13-24 months group, and 339 images for adult group. The adult group contained subjects between the ages of 19 to 66 years old.

4.2 Hypothesis 1

This hypothesis determined if there was a difference between image quality metrics of infants 0 to 6 months, 7 to 12 months, 13 to 24 months, and adults. A Welch's ANOVA was used to test if the means of the four groups were equivalent for each image quality metric, where H_0 denotes the null hypothesis and H_1 the alternative. The null hypothesis stated that the means for each group were equivalent and the alternative that at least one of the group's mean was different from the others. Or, more specifically:

$$H_0\colon \mu_{0to6} = \mu_{7to12} = \mu_{13to24} = \mu_{Adult}$$

$$H_1\colon \textit{Means are not all equal}$$

A Welch's ANOVA was used because the residuals did not appear to be homogenous and the Type I error is robust to non-homogenous variances (Liu, 2015). The residuals did not appear to be homogenous for any of the image quality metrics, making the Welch's ANOVA more appropriate. A Welch's ANOVA determined that at least one group's mean differed from the others. A post hoc comparison identified exactly which groups were statistically different and their respective confidence intervals. Like the Welch's ANOVA, the Games-Howell post hoc comparison of means did not assume equal sample sizes and homogeneity of variance.

4.2.1 Welch's ANOVA Results

A quantile-quantile plot (QQ) compares two distributions to each other (Marden, 2004). For all image quality metrics, a QQ plot was used to compare the residuals of the Welch's ANOVA to that of a normal distribution. If the residuals have a normal distribution, then the points in the QQ plot would form a straight line. Furthermore, a QQ plot can be used to detect potential outliers, heavy-tailed distributions, and skewness (Marden, 2004). The central limit theorem states that the distribution of a large random sample will converge to an approximately normal distribution, even if the real population is not normally distributed (Upton & Cook, 2008). Additionally, various studies have been conducted assessing the impact of non-normality of the Welch's ANOVA. The studies concluded that Type I and II error rates can be inflated by extreme non-normality such as an exponential distribution. However, when the group sizes were large, the residuals are heterogenous, and the residuals were approximately normal, even in cases of heavily tailed distributions, the Welch's ANOVA is quite robust.

The QQ plots for gray scale utilization, sharpness, pupil boundary circularity, and pupil to iris ratio revealed slightly skewed distributions. There was also some evidence of potential outliers in this study. The outliers were not caused from data collection errors and were believed to be representative of the population overall. Therefore, outliers were only removed if they were both a univariate and multivariate outlier. Univariate outliers were detected using a Grubbs outlier test, and multivariate outliers were detected using Mahalanobi's distance (Tabachnick & Fidell, 2013). A total of 22 outliers were removed, 5 from the 0 to 6 months group, 6 from the 7 to 12 months group, 8 from the 13 to 24 months group, and three from the adult group. The residual distributions for the image quality metrics were approximated as normal. The QQ plots and fitted vs residual plots for each image quality metric can be found in Appendix A. Margin adequacy and interlace had a score of 100 for all iris samples, regardless of group.

At a significance level of $\alpha = 0.01$, a significant difference was detected between groups for all image quality metrics. The effect size, ω^2, for the iris pupil concentricity and pupil boundary circularity was 0.01, indicating that the difference detected may not be practically significant. The group means, p-value, and effect size are reported in Table 4.1.

Table 4.1.
Welch's ANOVA Result Summary

	0to6	7to12	13to24	Adult	*p-value*	ω^2
	Mean					
Gray Scale Utilization	2.57	2.82	3.14	7.07	< 0.001	0.89
Iris Pupil Concentricity	97.32	97.4	97.21	97.21	0.002	0.01
Iris Pupil Contrast	65.54	74.85	78.57	66.66	< 0.001	0.13
Iris Radius	138.84	141.78	142.88	147.16	< 0.001	0.21
Iris Sclera Contrast	39.58	39.58	37.52	22.44	< 0.001	0.52
Pupil Boundary Circularity	95.63	96.03	95.42	96.09	0.004	0.01
Pupil to Iris Ratio	28.31	29.30	30.41	26.29	< 0.001	0.11
Scalar Quality	80.45	88.28	91.98	90.62	< 0.001	0.10
Sharpness	4.62	6.06	9.948	9.988	< 0.001	0.07
Usable Iris Area	79.60	81.15	81.07	85.35	< 0.001	0.04
Iris Detection Confidence	72.40	77.50	79.24	74.66	< 0.001	0.03

A Games-Howell post-hoc analysis was used to determine which group means differed for each image quality metric. The Games-Howell post-hoc analysis is summarized in Table 4.2, the bolded values highlight the groups where a significant difference was not detected, at a significance level of 0.01. Further examination of the metrics that had a small effect size for the Welch's ANOVA indicated that the pupil boundary circularity means were not statistically different for all the groups, but the pupil concentricity mean for infants 7 to 12 months old was statistically different from infants 13 to 24 months old and adults.

Table 4.2.

Games-Howell Post-Hoc Summary

Quality Metric	7to12-0to6		13to24-0to6		Adult-0to6		13to24-7to12		Adult-7to12		Adult-13to24	
	p	g*	p	g*	p	g*	p	g*	p	g*	p	g*
Gray Scale Utilization	<0.001	0.38	<0.001	0.90	<0.001	11.35	<0.001	0.47	<0.001	7.59	<0.001	7.15
Iris Pupil Concentricity	**0.642**	0.09	**.428**	-0.12	**0.499**	-0.12	0.004	-0.21	**0.012**	-0.22	**1.000**	0.00
Iris Pupil Contrast	<0.001	0.56	<0.001	0.90	**0.844**	0.08	<0.001	0.28	<0.001	-0.64	<0.001	-1.12
Iris Radius	<0.001	0.57	<0.001	0.80	<0.001	1.64	0.002	0.23	<0.001	1.11	<0.001	0.89
Iris Sclera Contrast	**1.00**	0.00	<0.001	-0.37	<0.001	-2.12	<0.001	-0.38	<0.001	-2.36	<0.001	-2.06
Pupil Boundary Circularity	**0.380**	0.13	**0.841**	-0.06	**0.288**	0.15	**0.015**	-0.19	**0.990**	0.02	**0.011**	0.21
Pupil to Iris Ratio	**0.099**	0.19	<0.001	0.44	<0.001	-0.50	0.001	0.24	<0.001	-0.74	<0.001	-1.10
Quality	<0.001	0.55	<0.001	0.96	<0.001	0.79	<0.001	0.36	0.005	0.22	**0.074**	-0.17
Sharpness	0.001	0.29	<0.001	0.55	<0.001	0.80	<0.001	0.44	<0.001	0.60	**1.000**	0.00
Usable Iris Area	**0.149**	0.17	**0.159**	0.17	<0.001	0.63	**0.999**	-0.01	<0.001	0.45	<0.001	0.48
Iris Detection Confidence	<0.001	0.37	<0.001	0.55	**0.253**	0.16	**0.104**	0.14	0.009	-0.22	<0.001	-0.39

Table 4.3 offers a better understanding how the groups differed. The letters denote which groups differed and which groups did not. If two groups share the same letter than no difference was detected between them.

Table 4.3.
Game-Howell Groupings

Group	Gray Scale Utilization	Iris Pupil Concentricity	Iris Pupil Contrast	Iris Radius	Iris Sclera Contrast	Pupil Boundary Circularity	Pupil to Iris Ratio	Quality	Sharpness	Usable Iris Area	Iris Detection Confidence
0 to 6	A	AB	A	A	A	A	A	A	A	A	A
7 to 12	B	A	B	B	A	A	A	B	B	A	B
13 to 24	C	B	C	C	B	A	B	C	C	A	B
Adult	D	AB	A	D	C	A	C	C	C	B	A

4.2.2 Hypothesis 1 Summary

In summary, there was at least one group that had a different mean for all the image quality metrics, except for pupil boundary circularity. A further investigation revealed that gray scale utilization and iris radius was the only quality metrics in which all groups differed. There was evidence that other image quality metrics were different for infants than adults such as pupil to iris ratio and usable iris area which are metrics that could be impacted by behavior or environment. Interestingly, iris radius was different for all four groups but improved with age which suggests that behavior might have played a crucial role in the differences detected. Other quality metrics that were impacted by behavior had this pattern observed as well. As this study did not record infant behavior the exact impact cannot be defined but there was substantial evidence that it played a role in the difference of image quality metrics for the infant groups. Scalar quality and sharpness were different for infants 0 to 6 months old and infants 7 to 12 months old but there was no difference detected between infants 13 to 24 months old and adults.

4.3 Hypothesis 2

This hypothesis tested if there was a difference in matching performance for infants and adults. Infants were split into three age groups 0 to 6 months, 7 to 12 months, and 13 to 24 months old. For the infants each age group was selected separately from the other groups, resulting in an infant’s iris that appeared in more than one age group. Because there were multiple visits in the infant data collection, an infant may have had irises from two separate visits that fell within the same age group. If an infant had an iris collected in two separate visits at an age where it fell within the same age group, the earliest visit was used. If an infant had only one unique iris sample in a particular visit, it was removed. Infants’ unique irises were also removed if they were not collected in at least two visits. No further cleaning was conducted for the adult irises. After the data were cleaned, the 0 to 6 months group had 29 unique irises with a total of 77 iris samples. The 7 to 12 months group had 141 samples from 52 unique irises, and the 13 to 24 months group had 162 samples from 57 unique irises. The adult iris group had 339 total samples from 113 unique irises.

By default, the iris camera used in this study had internal quality control measures to circumvent collecting poor quality irises which were not altered while acquiring the irises for both the infant and adult data collections. Quality control measures can be set while attempting to extract a template from an iris image. This hypothesis looked at iris images matching performance based on templates that were extracted with no quality control criteria in place. The scalar quality was determined by the Neurotechnology SDK. The templates used in this study were large and matching speed was slow – the Neurotechnology documentation recommended these settings to obtain the best matching accuracy. Figure 4.1 shows the DET curves for all four groups. The EERs reported were 1.54%, 0%, 0%, and 0% for the 0 to 6, 7 to 12, 13 to 24, and adult groups, respectively.

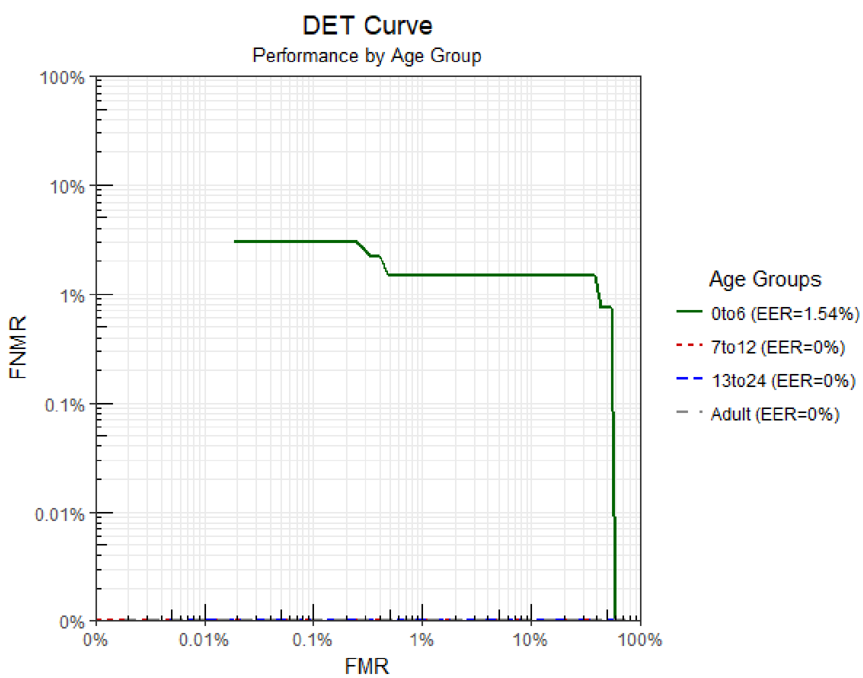

Figure 4.1. Performance by Age Group

Table 4.4 shows the FNMRs for all groups at different FMRs. Examining the overall performance of each group, infants 0 to 6 months old had the lowest performance with a FNMR of 3% at an FMR of 0.01%. At the same FMR, the FNMRs for the 7 to 12, 13 to 24, and adult groups were 0%, 0%, and 0%, respectively.

Table 4.4.
FNMRs by Age Group

FMR	0to6	7to12	13to24	Adult
0.01%	2.99%	0%	0%	0.6%
0.10%	2.99%	0%	0%	0.6%
1%	1.49%	0%	0%	0.6%
10%	1.49%	0%	0%	0.6%

The performance results indicated that iris recognition for infants and adults were similar. However, the youngest age group, infants 0 to 6 months old, did have slightly worse performance and in fact, the only group that did not have a FNMR of 0% at a FMR of 0.01%, 0.1%, 1%, and 10%.

4.3.1 Hypothesis 2 Summary

In summary, there was no difference between infant and adult iris recognition performance, except for infants 0 to 6 months old. However, at an FMR of 0.01% the FNMR of the 0 to 6 months old age group was 3%. Obviously, the desired FMR or FNMR is operational scenario driven, but in many scenarios this performance should be adequate, regardless of age. For example, in IREX IV the Neurotechnology SDK had a FNMR of 3% and 4% at an FMR of 0.01% for enrolled population sizes of 10,000 and 1,600,000, respectively (Quinn, Grother, & Ngan, 2013). Additionally, the report mentions that accuracy is less dependent on enrolled population size than other biometric modalities and that the number of enrolled users can be increased without inflating false match and non-match rates (Quinn et al., 2013). Another important conclusion from this hypothesis is that scalar quality was an adequate predictor for performance for both adult and infant iris recognition.

4.4 Hypothesis 3

This hypothesis evaluated if genuine similarity scores change over time using a mixed linear regression model. The mixed linear regression represented in Equation 6 models the genuine similarity score, s_{ij}, for the *i*-th eye and the *j*-th score of the iris camera:

$$s_{ij} = \beta_0 + \beta_{\Delta T} \Delta T_{ij} + \beta_{\Delta D} \Delta D_{ij} + \beta_S S_{ij} + \beta_A A_{ij} + b_{i0} + b_{\Delta T} \Delta T_{ij} + b_{i\Delta D} \Delta D_{ij} + \beta_S S_{ij} + b_{iA} A_{ij} + e_{ij} \tag{6}$$

Where, β_k represents the fixed effects of the k-th predictor variable and b_{ik} denotes a random effect of the i-th eye for the k-th predictor variable; the elapsed time in 30-day increments between when two iris samples were captured is denoted as ΔT; Dilation differences, ΔD, is a measure of the differences between the pupil to iris ratio of two iris images; S, is the smallest sharpness value between two images; A, is the smallest usable iris area between two images; e_{ij} are the residuals. These covariates were selected based on their parsimony and Bayesian information criterion, which optimizes the model complexity and model's ability to fit the data (Upton & Cook, 2008). This hypothesis only evaluated infant irises that were acquired in more than one visit and had more than one unique iris acquired for a particular visit.

This research was particularly interested in the fixed effect predictor elapsed time (in 30-day intervals), which represents the population average rate of change in similarity scores over time. It is important to note that this rate of change cannot be generalized to another set of infants, a replication will need to be conducted to strengthen this research and understanding of how infants' genuine similarity scores change over time for iris recognition. As seen in Figure 4.2, each unique eye seemed to have a different similarity score rate of change over time. Most irises observed a downward trend in similarity scores over time, whereas, some subjects had an increase in similarity scores such as 02333LE, 02040RE, 02023RE, 01982LE.

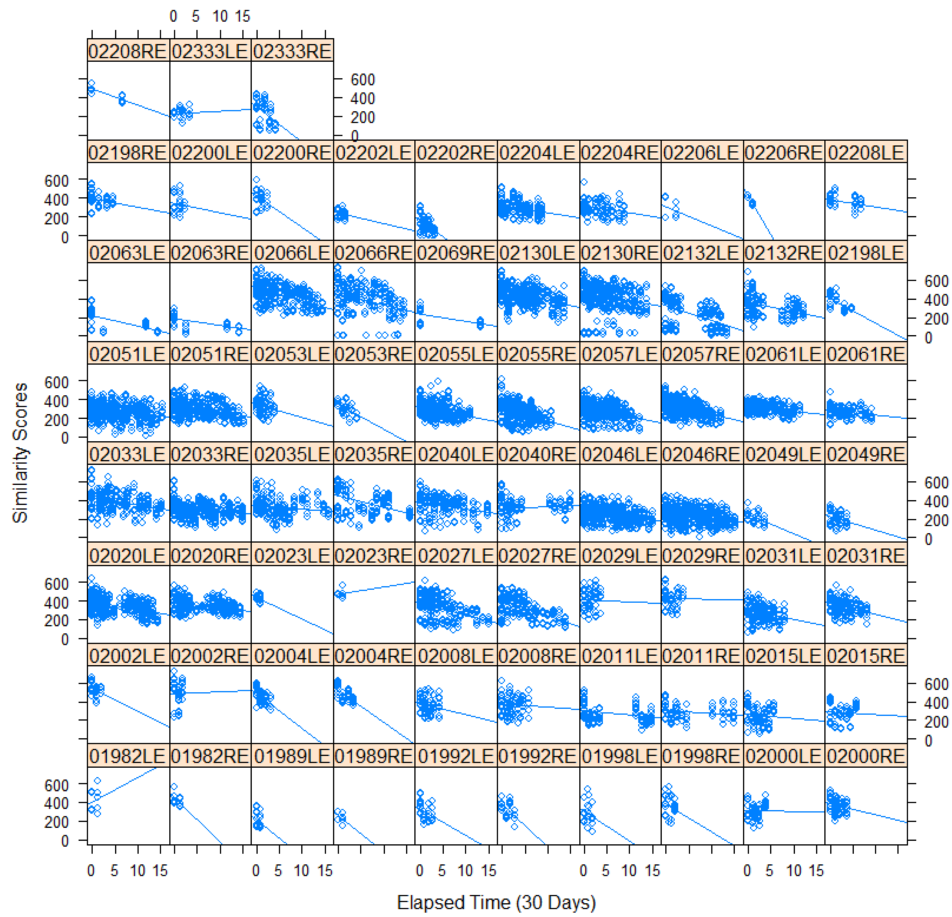

Figure 4.2. Subject Specific Similarity Scores Over Time

The regression coefficients for the fixed effects of the iris camera are shown in Table 4.5. Each fixed effect shows the average population rate of change in genuine similarity score for a given covariate. The p-values were given by a t-test, where the null hypothesis $\beta_k = 0$ versus the alternative, $\beta_k \neq 0$, given all the other covariates in the model. At an $\alpha = 0.01$ and the p-value < 0.001, the null hypothesis was rejected in favor of the alternative, that given all the other predictors in the model $\beta_{\Delta T}$ is not equal to zero and that the average rate of change in genuine similarity score over time was significant.

Table 4.5.
Fixed Effect Coefficients

Coefficient	Fixed Effect	*p-value*
Intercept	-227.93	< 0.001
$\beta_{\Delta T}$	-5.16	< 0.001
β_{S}	58.50	< 0.001
β_{A}	6.40	< 0.001
$\beta_{\Delta D}$	-5.22	< 0.001

The random effects' standard deviation across the subjects are listed in Table 4.6. A variance of zero would indicate that the corresponding fixed effects alone were able to fit all of the subjects perfectly. As shown, none of the standard deviations of the random effects is zero suggesting that the random effects were relevant to the model. The random effects implied that each unique iris had its own rate of change for each covariate in the model.

Table 4.6.
Random Effect Standard Deviations

Coefficient	SD
Intercept	108.88
$\beta_{\Delta T}$	2.05
β_{S}	43.92
β_{A}	1.46
$\beta_{\Delta D}$	2.01

Regression diagnostics showed that the residuals were homogenous and normally distributed. Given that $\beta_{\Delta T} \neq 0$, there is an evident downward trend per 30 days' lapse in time. The $\beta_{\Delta T}$ coefficient indicated that for every 30 days' lapse that, on average, the similarity score will decrease by approximately a score 5. Within approximately one year, the similarity score can be expected to drop, on average, by 60. Again, this rate of change only serves as a first step in what should be a replication of this analysis on another group of infants, but the outcome is encouraging because a change of 5 in a genuine similarity score is not a large change.

4.4.1 Hypothesis 3 Summary

The average rate of change of the genuine similarity scores over elapsed time was statistically significant for the iris camera, however, the observed change over time does not appear to be practically significant. The average rate of change was a decrease of 5 in genuine similarity scores for every elapsed 30 days. It is important to note that this rate of change at this time only applies to this particular sample of infants. This analysis would have to be replicated on a separate sample of infants before any generalizations to the population can be made.

The most significant finding of this hypothesis was that the biggest impact on performance was not the time between samples but the change in dilation, the difference of sharpness between two images, and the amount of usable area in an iris image. The outcomes of this study agree and support the conclusions and results from IREX 6 report (Grother et al., 2015). Except that, in this study the subjects were infants that were uncooperative and resulted in varying degrees of poor quality images.

CHAPTER 5. CONCLUSIONS AND FUTURE WORK

This chapter is divided into two sections. The first section makes inferences and conclusions about the three hypotheses examined in this study. The second section outlines future work to be done in infant iris recognition, including recommendations based on this study's findings, and recommendations based on what this study has not covered.

5.1 Conclusions

Many conclusions can be drawn from this study: the first, image quality metrics such as, usable iris area, sharpness, dilation, and iris radius were impacted by a subject's age. These metrics can also be affected by behavior. For example, there was a clear difference between all groups' iris radii. As age increased the iris radii increased and the values were more consistent. It is known that from birth to about two years old infants undergo a major transformation in their attentive and visual ability, which would indicate that as an infant gets older they would be more cooperative. For example, with sharpness, which measures the degree of blur in an iris image, there were differences observed for infants 0 to 12 months compared infants 13 to 24 months old and adults. The average sharpness score was higher for infants 13 to 24 months old and adults than infants 0 to 12 months (the 0 to 6 months and 7 to 12 months groups were not different), indicating that the amount of blur in an iris image improved as a subject gets older. Again, pupil to iris ratio was different for infants 0 to 12 months old compared to adults or infants 13 to 24. The pupil to iris ratio variance decreased as age increased, indicating that cooperation improved with age. Finally, usable iris area was different for infants and adults. This metric could be affected by behavior or environment, because the collection environment was different for the adult and infant collections, it is hard to infer if the differences were attributed to the device, environment, or a combination of both factors.

There were many challenges with infant cooperativeness during data collection. The general interaction for the iris camera was the same, regardless if a subject was an infant or adult. The infants most certainly behaved differently during acquisition. The infants were held by their parents during this interaction, but the infants were sometimes crying, moving, or looking away. The parents themselves could have moved. In some cases, parents had to stand in various places because of the way they held the infants, the parent's height, or a combination of both. As this was a secondary analysis, in the infant collection, many devices were used, and the subjects had to interact with several devices which may have overstimulated them. The iris collection was always the last modality to collect, therefore if the infant was sleeping, they were not awoken for collection, and if they just awoken from a nap and were sometimes irritable.

The second conclusion, in general, infants 0 to 6 months old had worse recognition performance than infants older than 6 months old and adults. After 6 months old, the performance was the same as adult iris recognition. At FMR of 0.01%, infants 0 to 6 months old had a FNMR of 3% where the other groups had a FNMR of 0%. It is important to note that a FNMR of 3% is not bad. For example, in the IREX IV report, the Neurotechnology SDK had a FNMR of 3% and 4% for adults in an operation scenario and enrollment populations of 10,000 and 1,600,00, respectively (Quinn et al., 2013). These are promising results for infants of all ages, because iris recognition is known to scale well without increasing error rates substantially (Quinn et al., 2013).

Scalar image quality was a good predictor of performance. The image quality assessment algorithm appeared to work properly for the iris camera. Performance of infant iris recognition may be susceptible to an infant's behavior, however, no adjustments to iris recognition algorithms, quality assessment algorithms, or the iris images themselves was necessary. Because there were differences detected which appeared to be caused by age, testing infant iris recognition on robust cameras meant to capture irises in non-ideal situations may be beneficial – specifically devices that are used for iris recognition on the move or that allow for discrete capture of iris without a subject's participation or knowledge.

There has only been one other infant iris study that evaluated matching performance. Corby et al. (2006) studied iris recognition in infants between the ages of 1.5 to 8 years old. The performance of this study was evaluated a year after enrollment. About 99.4% of the infants between the ages of 1.5 to 8 years old were correctly identified. In this study, infants at a younger age had similar performance. For example, infants 7 to 24 months had a FNMR of 0% at a FMR of 0.01%, and infants 0 to 6 months old had a FNMR of 3%.

Finally, the biggest impact on performance was not the time between samples but the change in dilation, the difference of sharpness between two images, and the amount of usable area in an iris image.

5.2 Future Work

The first recommendation for future work is to collect more infant iris data longitudinally and on a wider selection of iris cameras. The work in this study is the first publicly available research that extensively examined infant iris recognition performance longitudinally for infants 0 to 24 months old. Replication of this research and the methods used will support and aid in furthering infant iris recognition research.

This study did not record an infant's behavior or interactions with an iris recognition camera. Doing so would provide a strong understanding of an infant's behavior and a certain behavior's impact on performance.

A comparison of the same subjects across different biometric modalities will help the biometric research community understand the most suitable biometrics for infants. One important question that remains unanswered is whether an infant's physical iris pattern changes over time. All though the insignificant decrease in biometric performance is a strong indicator that is not the case, a further investigation is warranted.

All the data in this study was collected in a controlled lab environment. A further investigation of infant iris recognition in unconstrained environments may have significant impact on understanding the practical uses of infant iris recognition for identifying infants in healthcare, police, vaccination coverage, or homeland security applications.

A major challenge in this study was to compare performance and image quality results to other infant biometric studies. For example, one study defines image quality as good or poor and another defined image quality as failed enrollment, partial enrollment, and marginal enrollment. Without knowing exactly what "good" or "partial enrollment" means a proper comparison of results is difficult. These same challenges can also be seen in current biometric definitions, such as acceptable biometric capture attempt or quality. These definitions are well known for adults but not so much for infants. Updating or re-defining biometric definitions to include infants will create a common language for future infant biometric studies and will simplify comparisons and references to other studies.

Finally, it is important to put in place best practices for collecting iris samples from infants. A best practices document will help guide future research studies, improve quality of samples, and really should be considered for all biometric modalities of infants. Test or lab administrators may be able to provide helpful insight for improving image quality and strengthening the biometric communities understanding of the challenges of having infants as test subjects. This could also lead to iris cameras or other biometric systems designed specifically for infants.

APPENDIX A. HYPOTHESIS 1 DIAGNOSTIC PLOTS

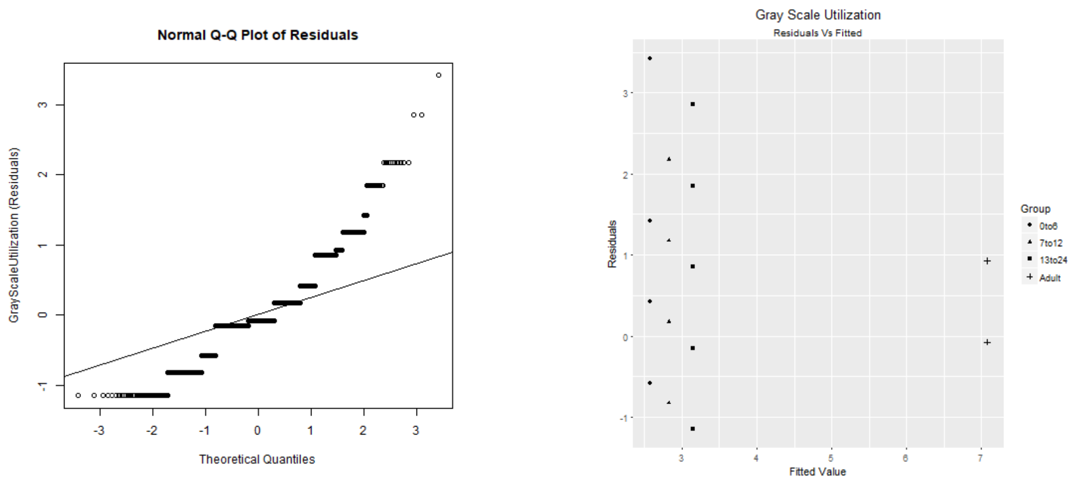

Figure A.1. Gray scale utilization diagnostic plots

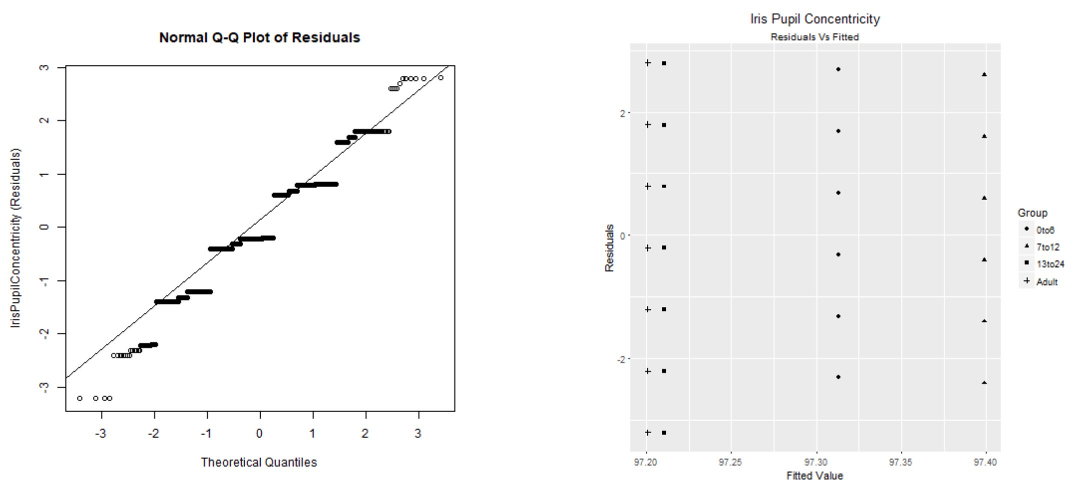

Figure A.2. Iris pupil concentricity diagnostic plots

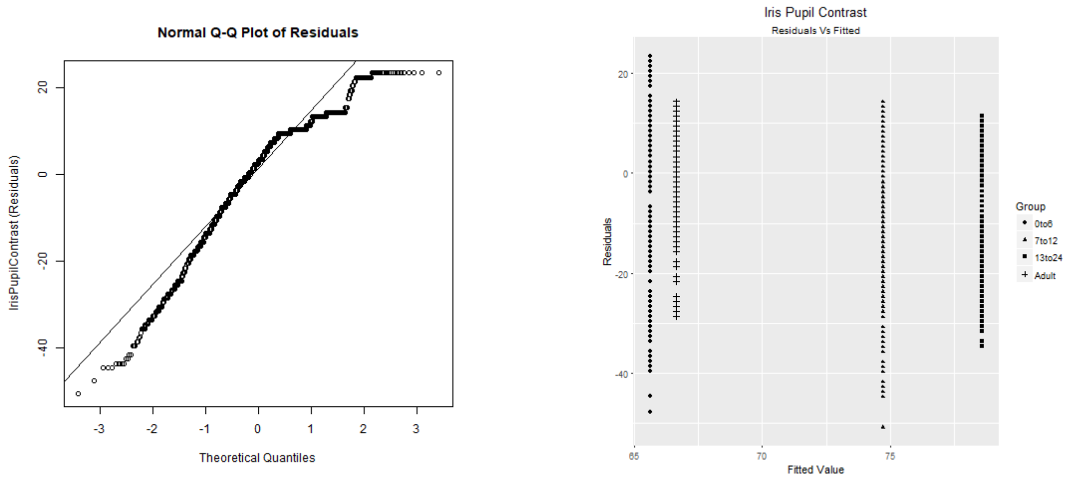

Figure A.3. Iris pupil contrast diagnostic plots

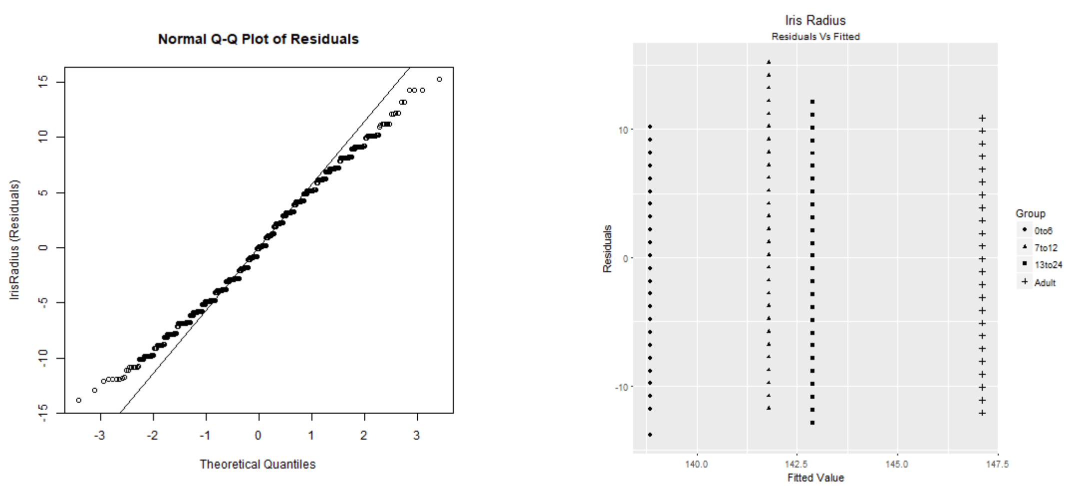

Figure A.4. Iris radius diagnostic plots

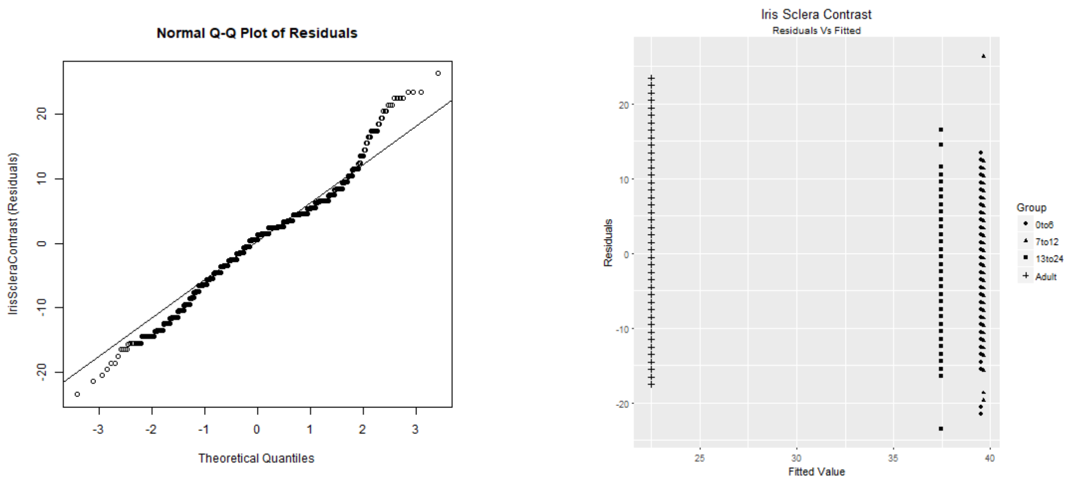

Figure A.5. Iris sclera contrast diagnostic plots

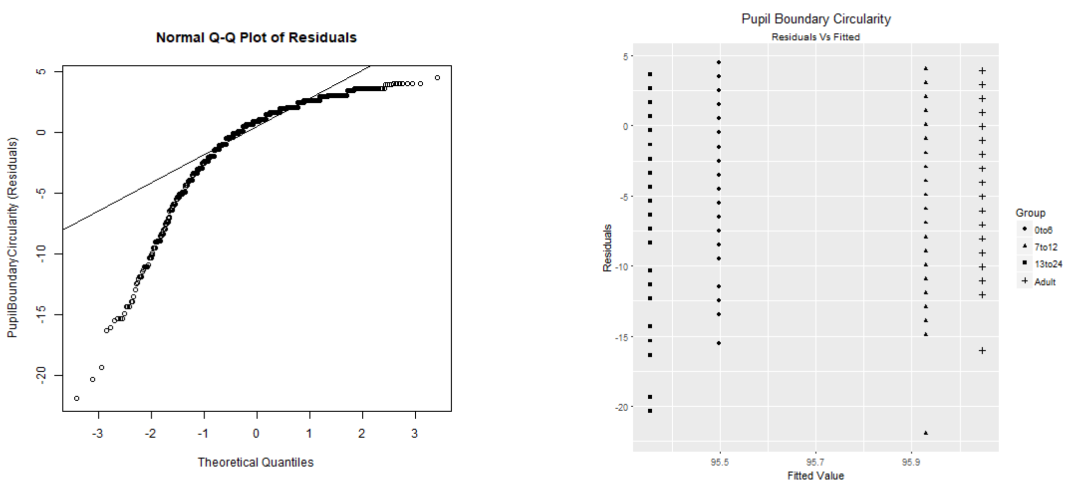

Figure A.6. Pupil boundary circularity diagnostic plots

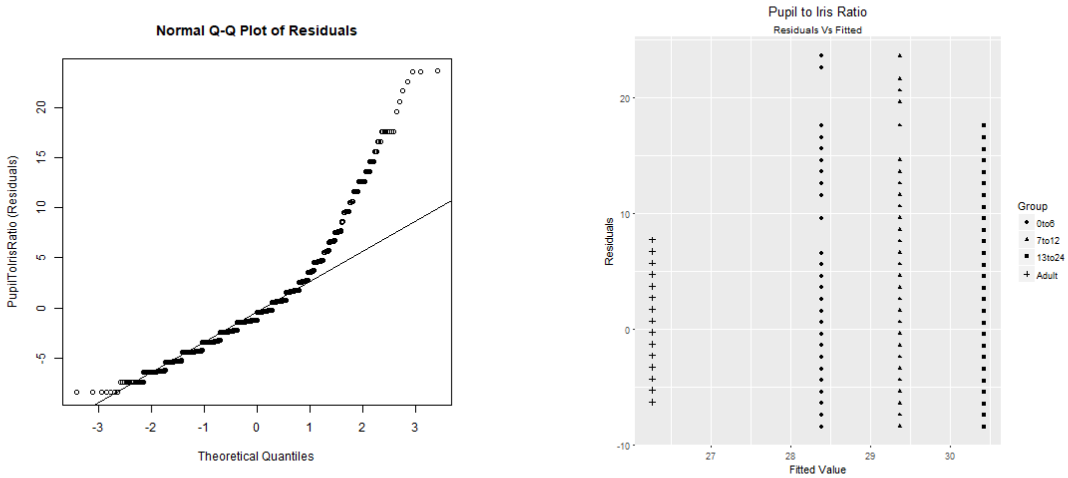

Figure A.7. Pupil to iris ratio diagnostic plots

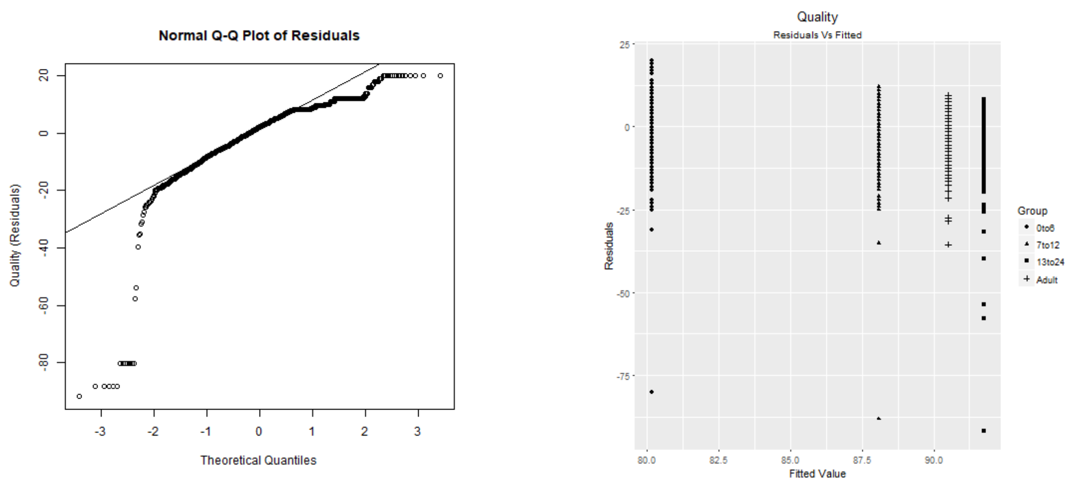

Figure A.8. Scalar quality diagnostic plots

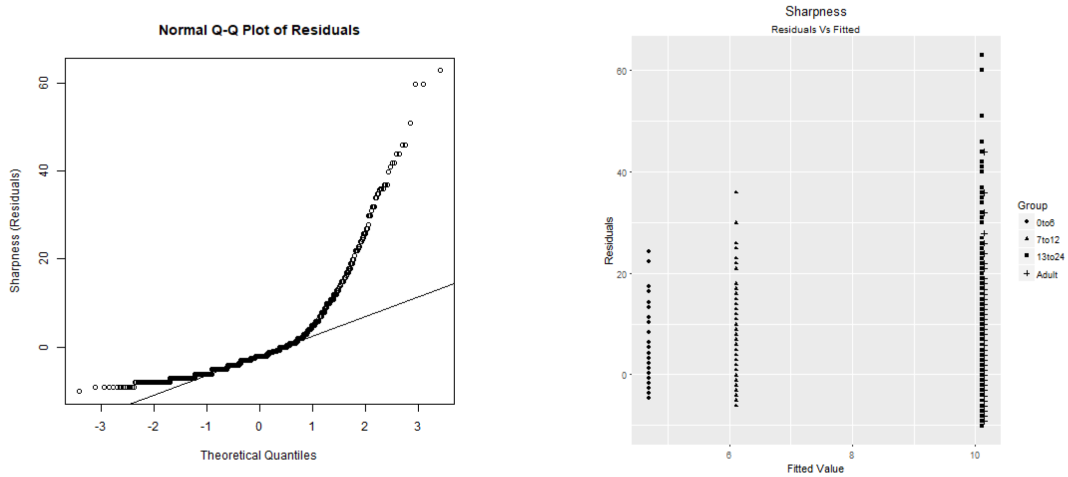

Figure A.9. Sharpness diagnostic plots

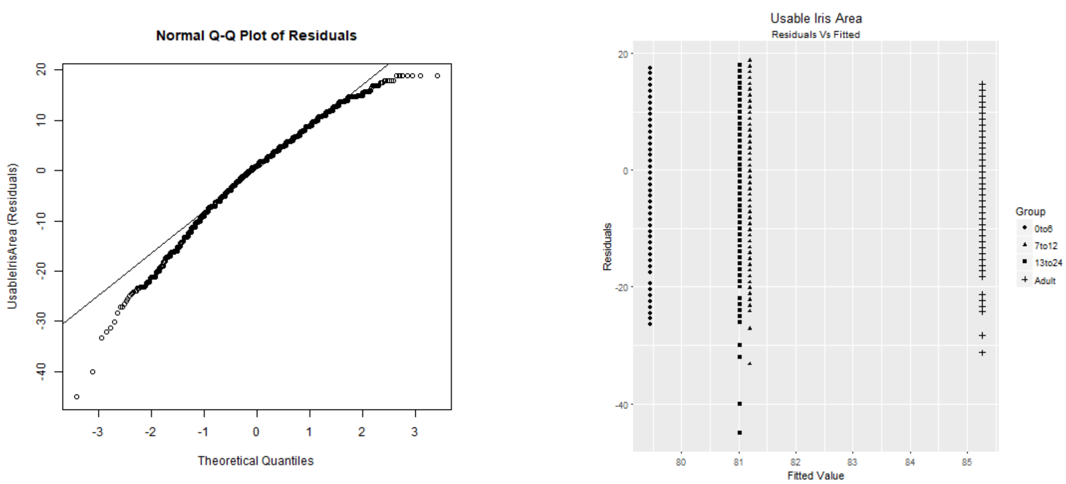

Figure A.10. Usable iris area diagnostic plots

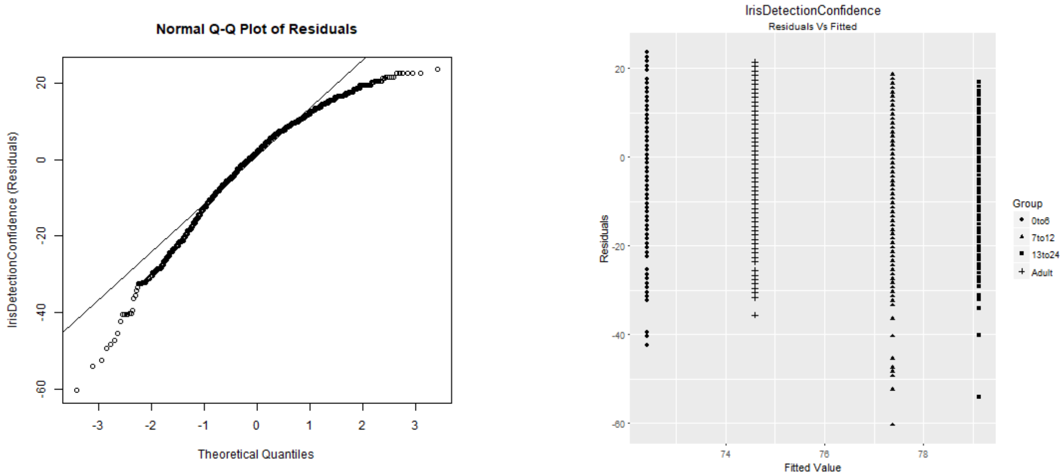

Figure A.11. Iris detection confidence diagnostic plots

REFERENCES

Ackerman, D. (2016). Optics of Iris Imaging Systems. In K. W. Bowyer & M. J. Burge (Eds.), *Handbook of Iris Recognition* (pp. 63–88). London: Springer London. https://doi.org/10.1007/978-1-4471-6784-6_3

AOptix. (2011). Insight Duo. Retrieved from www.bordotek.com/Download/aoptix/Aoptix-InSight-Duo-Datasheet.pdf

Bachenheimer, D. (2016). *UNHCR Biometric Identity Management System*. Conference presented at the International Biometric Performance Conference, Gaithersburg, MD.

Baker, S. E., Bowyer, K. W., & Flynn, P. J. (2009). Empirical Evidence for Correct Iris Match Score Degradation with Increased Time-Lapse between Gallery and Probe Matches. In M. Tistarelli & M. S. Nixon (Eds.), *Advances in Biometrics* (Vol. 5558, pp. 1170–1179). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-01793-3_118

Barra, S., Casanova, A., De Marsico, M., & Riccio, D. (2014). Babies: Biometric authentication of newborn identities by means of ear signatures (pp. 1–7). IEEE. https://doi.org/10.1109/BIOMS.2014.6951528

Bergmuller, T., Debiasi, L., Uhl, A., & Sun, Z. (2014). Impact of sensor ageing on iris recognition (pp. 1–8). IEEE. https://doi.org/10.1109/BTAS.2014.6996274

Bertillon, A. (1886). *La couleur de l'Iris*. Masson.

Bharadwaj, S., Bhatt, H. S., Singh, R., Vatsa, M., & Singh, S. K. (2010). Face recognition for newborns: A preliminary study (pp. 1–6). IEEE. https://doi.org/10.1109/BTAS.2010.5634500

Bridges, C. D. (2015). *Biology 201/203: Human anatomy and physiology*. Plymouth, MI: Hayden-McNeil.

Clarke, R. (1994). Human Identification in Information Systems: Management Challenges and Public Policy Issues. *Information Technology & People*, *7*(4), 6–37. https://doi.org/10.1108/09593849410076799

Cole, J. (2016, May). *Expanding the role of analytics in child exploitation cases*. Presented at the International Biometrics Performance Conference, Gaithersburg, MD.

Corby, P. M., Schleyer, T., Spallek, H., Hart, T. C., Weyant, R. J., Corby, A. L., & Bretz, W. A. (2006). Using Biometrics for Participant Identification in a Research Study: A Case Report. *Journal of the American Medical Informatics Association*, *13*(2), 233–235. https://doi.org/10.1197/jamia.M1793

Daugman, J. (1994). *U.S. Patent No. 5,291,560*.

Daugman, J. (2004). How Iris Recognition Works. *IEEE Transactions on Circuits and Systems for Video Technology*, *14*(1), 21–30. https://doi.org/10.1109/TCSVT.2003.818350

Du, Y. (2006). Review of iris recognition: cameras, systems, and their applications. *Sensor Review*, *26*(1), 66–69. https://doi.org/10.1108/02602280610640706

Dunstone, T., & Yager, N. (2009). *Biometric system and data analysis: design, evaluation, and data mining*. New York: Springer.

Dutch Ministry of the Interior and Kingdom Relations. (2005). *Evaluation Report Biometrics Trial 2b or not 2b* (Technical Report). Dutch Ministry of the Interior and Kingdom Relations.

Ebay. (2018). Silicon Valley Compucycle Ebay Store. Retrieved from https://www.ebay.com/itm/NEW-Panasonic-Authenticam-BM-ET100US-Iris-Recognition-Camera-Security-Scanner-/112461862219

Farlex. (2018). gestation. *The Free Dictionary*. Retrieved from https://medical-dictionary.thefreedictionary.com/gestation

Fenker, S. P., & Bowyer, K. W. (2011). Experimental evidence of a template aging effect in iris biometrics (pp. 232–239). IEEE. https://doi.org/10.1109/WACV.2011.5711508

Flom, L., & Safir, A. (1987). *U.S. Patent No. 4,641,349*. Retrieved from http://www.google.com/patents/US4641349

Fulcrum Biometrics LLC. (2018). IriShield-USB BK2121U Auto-Capture Iris Camera. Retrieved January 11, 2018, from http://www.fulcrumbiometrics.com//ProductDetails.asp?ProductCode=102152

Global Good Intellectual Ventures. (2017). Global Health Technologies: Invention by saving lives. Retrieved March 26, 2018, from http://www.intellectualventures.com/landing_pages/globalgood/assets/files/Brochure%20-%20Global%20Health%20Technologies%20v1.1.pdf

Grother, P., Matey, J. R., & Quinn, G. W. (2015). IREX VI: mixed-effects longitudinal models for iris ageing response to Bowyer and Ortiz. *IET Biometrics*, *4*(4), 200–205. https://doi.org/10.1049/iet-bmt.2015.0043

Grother, P., Matey, J., Tabassi, E., Quinn, G., & Chumakov, M. (2013). *IREX VI: Temporal stability of iris recognition accuracy* (Technical Report No. 7948) (p. 67). NIST.

Grother, P., Quinn, G. W., Matey, J. R., Ngan, M., Salamon, W., Fiumara, G., & Watson, C. (2012). *IREX III: Performance of iris identification algorithms* (Technical Report No. 7836) (p. 102). NIST.

Harvey, J., Campbell, J., Elliott, S., Brockly, M., & Adler, A. (2017). Biometric Permanence: Definition and Robust Calculation. In *Systems Conference (SysCon), 2017 Annual IEEE International* (pp. 1–7). IEEE.

Jain, A., Arora, S., Best-Rowden, L., Cao, K., Sudhish, P. S., & Bhatnagar, A. (2015). *Biometrics for Child Vaccination and Welfare: Persistence of Fingerprint Recognition for Infants and Toddlers* (Technical Report No. MSU-CSE-15-7) (p. 17). Michigan State University.

Jain, A., Arora, S., Cao, K., Best-Rowden, L., & Bhatnagar, A. (2017). Fingerprint Recognition of Young Children. *IEEE Transactions on Information Forensics and Security*, *12*(7), 1501–1514. https://doi.org/10.1109/TIFS.2016.2639346

Jain, A., Arora, S. S., Best-Rowden, L., Cao, K., Sudhish, P. S., Bhatnagar, A., & Koda, Y. (2016). Giving Infants an Identity: Fingerprint Sensing and Recognition (pp. 1–4). ACM Press. https://doi.org/10.1145/2909609.2909612

Jain, A., Cao, K., & Arora, S. S. (2014). Recognizing infants and toddlers using fingerprints: Increasing the vaccination coverage (pp. 1–8). IEEE. https://doi.org/10.1109/BTAS.2014.6996252

Jain, A., Ross, A. A., & Nandakumar, K. (2011). *Introduction to biometrics*. New York: Springer.

Jain, A., Ross, A., & Prabhakar, S. (2004). An Introduction to Biometric Recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, *14*(1), 4–20. https://doi.org/10.1109/TCSVT.2003.818349

Jia, W., Cai, H.-Y., Gui, J., Hu, R.-X., Lei, Y.-K., & Wang, X.-F. (2012). Newborn footprint recognition using orientation feature. *Neural Computing and Applications*, *21*(8), 1855–1863. https://doi.org/10.1007/s00521-011-0530-9

Jia, W., Hu, R.-X., Gui, J., & Lei, Y.-K. (2010). Newborn Footprint Recognition Using Band-Limited Phase-Only Correlation. In D. Zhang & M. Sonka (Eds.), *Medical Biometrics* (Vol. 6165, pp. 83–93). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-13923-9_9

Jiang, J. (2007). *Linear and generalized linear mixed models and their applications*. New York ; London: Springer.

Jillela, R., & Ross, A. A. (2016). Methods for Iris Segmentation. In K. W. Bowyer & M. J. Burge (Eds.), *Handbook of Iris Recognition* (pp. 137–184). London: Springer London. https://doi.org/10.1007/978-1-4471-6784-6_7

Johnston, R. (1991). *Can Iris Patterns Be Used to Identify People?* (No. LA-12331-PR) (pp. 81–86). Chemical and Laser Sciences Division.

Joun, S., Kim, H., Chung, Y., & Ahn, D. (2003). An Experimental Study on Measuring Image Quality of Infant Fingerprints (pp. 1261–1269). Presented at the International Conference on Knowledge-Based and Intelligent Information & Engineering Systems.

JTC 1/ SC 37. (2006). ISO/IEC 19795 Information technology - Biometric performance testing and reporting - Part 1: Principles and framework.

JTC 1/SC 37. (2011). *ISO/IEC 29794-6 Information technology-Biometric sample quality - Part 6: Iris image data* (International standard edition). JTC 1/SC 37.

JTC 1/SC 37. (2012). *ISO/IEC 29794-1 Information technology-Biometric sample quality - Part 1: Framework* (International standard edition). JTC 1/SC 37.

JTC 1/SC 37. (2017). *ISO/IEC 2382-37 Information technology - Vocabulary - Part 37: Biometrics*.

Kotzerke, J., Arakala, A., Davis, S., Horadam, K., & McVernon, J. (2014). Ballprints as an infant biometric: A first approach (pp. 36–43). IEEE. https://doi.org/10.1109/BIOMS.2014.6951533

Kotzerke, J., Davis, S., Horadam, K., & McVernon, J. (2013). Newborn and infant footprint crease pattern extraction (pp. 4181–4185). IEEE. https://doi.org/10.1109/ICIP.2013.6738861

Lemes, R. P., Bellon, O. R. P., Silva, L., & Jain, A. K. (2011). Biometric recognition of newborns: Identification using palmprints (pp. 1–6). IEEE. https://doi.org/10.1109/IJCB.2011.6117475

Liu, H. (2015). Comparing Welch's ANOVA, a Kruskal-Wallis test and traditional ANOVA in case of Heterogeneity of Variance, 48.

Madhu, M., Moorthi, M., Sathishkumar, S., & Vimalagopalan, T. S. (2017). A novel approach for new-born authentication and verification using multibiometric features. *Biomedical Research, 28*(5). Retrieved from http://www.alliedacademies.org/articles/a-novel-approach-for-newborn-authentication-and-verification-using-multibiometric-features.html

Maltoni, D., Maio, D., Jain, A., & Prabhakar, S. (Eds.). (2009). *Handbook of fingerprint recognition* (2nd ed). London: Springer.

Mansfield, A., & Wayman, J. (2002). *Best Practices in Testing and Reporting Performance of Biometric Devices ver 2.01*. Centre for Mathematics and Scientific Computing.

Marden, J. I. (2004). Positions and QQ Plots. *Statistical Science*, *19*(4), 606–614. https://doi.org/10.1214/088342304000000512

Matey, J. R., Naroditsky, O., Hanna, K., Kolczynski, R., LoIacono, D. J., Mangru, S., … Zhao, W. Y. (2006). Iris on the Move: Acquisition of Images for Iris Recognition in Less Constrained

Environments. *Proceedings of the IEEE*, *94*(11), 1936–1947. https://doi.org/10.1109/JPROC.2006.884091

Mehrotra, H., Vatsa, M., Singh, R., & Majhi, B. (2013). Does Iris Change Over Time? *PLoS ONE*, *8*(11), e78333. https://doi.org/10.1371/journal.pone.0078333

Neurotechnology. (2017). *Neurotechnology developer's guide 10.0*.

O'Connor, K. (2013). *Examination of stability in fingerprint recognition across force levels* (PhD Thesis). Purdue University.

Online Medical Dictionary. (2018a). Child - Online Medical Dictionary. Retrieved January 17, 2018, from www.online-medical-dictionary.org/definitions-c/child.html

Online Medical Dictionary. (2018b). Infant - Online Medical Dictionary. Retrieved January 17, 2018, from http://www.online-medical-dictionary.org/definitions-i/infant.html

Oxford Dictionary. (2018a). Definition of principal component analysis. Retrieved March 26, 2018, from https://en.oxforddictionaries.com/definition/principal_component_analysis

Oxford Dictionary. (2018b). ground truth. Retrieved from https://en.oxforddictionaries.com/definition/ground_truth

Oyster, C. W. (1999). *The human eye: structure and function*. Sunderland, Mass: Sinauer Associates.

Petry, B. (2015). *The stability of the iris as a biometric modality* (Masters). Purdue University.

Quinn, G. W., Grother, P., & Ngan, M. (2013). *IREX IV part 1 : evaluation of iris identification algorithms* (No. NIST IR 7949). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.7949

Remington, L. A. (2005). *Clinical anatomy of the visual system*. St. Louis, Mo.: Elsevier. Retrieved from http://www.sciencedirect.com/science/book/9780750674904

Rhcastilhos. (2018). Schematic diagram of the human eye with annotations. Retrieved from https://commons.wikimedia.org/wiki/File:Schematic_diagram_of_the_human_eye_en.svg#/media/File:Schematic_diagram_of_the_human_eye_en.svg

Roy, D. A., & Soni, U. S. (2016). Iris segmentation using Daughman's method (pp. 2668–2676). IEEE. https://doi.org/10.1109/ICEEOT.2016.7755178

Sazonova, N., Hua, F., Liu, X., Remus, J., Ross, A., Hornak, L., & Schuckers, S. (2012). A study on quality-adjusted impact of time lapse on iris recognition (pp. 83711W-83711W – 9). https://doi.org/10.1117/12.919642

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed). Boston: Pearson Education.

Tabassi, E., Grother, P., & Salamon, W. (2011). *IREX II iris quality calibration and evaluation: Performance of iris image quality assessment algorithms* (Technical Report No. 7820) (p. 226). NIST.

Tabassi, E., Wilson, C., & Watson, C. (2004). *NISTIR 7151 Fingerprint Image Quality* (p. 72).

Thainimit, S., Alexandre, L. A., & de Almeida, V. M. N. (2013). Iris surface deformation and normalization (pp. 501–506). IEEE. https://doi.org/10.1109/ISCIT.2013.6645910

Tiwari, S., Kumar, S., Kumar, S., Sinha, G., Delhi Section, & INDIAcom (Conference). (2015). Ear Recognition for Newborns. In *International Conference on Computing for Sustainable Global Development*. Retrieved from http://ieeexplore.ieee.org/servlet/opac?punumber=7088896

Tiwari, S., Singh, A., & Singh, S. (2011). Newborn's ear recognition: Can it be done? In *2011 International Conference on Image Information Processing* (pp. 1–6). https://doi.org/10.1109/ICIIP.2011.6108944

Tiwari, S., Singh, A., & Singh, S. (2012a). Can face and soft-biometric traits assist in recognition of newborn? (pp. 74–79). IEEE. https://doi.org/10.1109/RAIT.2012.6194483

Tiwari, S., Singh, A., & Singh, S. (2012b). Comparison of Adult and Infant Ear Images for Biometric Recognition. *International Journal of Biometrics and Bioinformatics*, *6*(6).

Tiwari, S., Singh, A., & Singh, S. (2013). Multimodal Database of Newborns for Biometric Recognition. *International Journal of Bio-Science and Bio-Technology*, *5*, 89–99.

Tiwari, S., Singh, A., & Singh, S. K. (2012c). Can Ear and Soft-Biometric Traits Assist in Recognition of Newborn? In D. C. Wyld, J. Zizka, & D. Nagamalai (Eds.), *Advances in Computer Science, Engineering & Applications* (Vol. 166, pp. 179–192). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30157-5_19

Tiwari, S., & Singh, S. (2012). Face recognition for newborns. *IET Biometrics*, *1*(4), 200–208. https://doi.org/10.1049/iet-bmt.2012.0040

Tome-Gonzalez, P., Alonso-Fernandez, F., & Ortega-Garcia, J. (2008). On the Effects of Time Variability in Iris Recognition (pp. 1–6). IEEE. https://doi.org/10.1109/BTAS.2008.4699329

Trokielewicz, M. (2015). Linear regression analysis of template aging in iris biometrics (pp. 1–6). IEEE. https://doi.org/10.1109/IWBF.2015.7110233

Upton, G. J. G., & Cook, I. (2008). *A dictionary of statistics* (2nd ed., rev). Oxford ; New York: Oxford University Press.

Watson, C., Fiumara, G., Tabassi, E., Salamon, W., & Flanagan, P. (2014). *Fingerprint Vendor Technology Evaluation* (No. NIST IR 8034). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8034

Wayman, J. (2000). *National biometric test center: Collected works 1997-2000* (Biometric Consortium of the US Government interest group on biometric authentication). San Jose, CA: San Jose State University.

Weingaertner, D., Bellon, O. R. P., Silva, L., & Cat, M. N. (2008). Newborn's biometric identification: Can it be done? In *VISAPP (1)* (pp. 200–205). Retrieved from https://www.researchgate.net/profile/Luciano_Silva10/publication/221415455_Newborn's_Biometric_Identification_Can_it_be_done/links/0c96052b8327747805000000.pdf

Wu, C., Tulyakov, S., & Govindaraju, V. (2006). Image quality measures for fingerprint image enhancement. In *International Workshop on Multimedia Content Representation, Classification and Security* (pp. 215–222). Springer.

Yuan, X., & Shi, P. (2005). A Non-linear Normalization Model for Iris Recognition. In S. Z. Li, Z. Sun, T. Tan, S. Pankanti, G. Chollet, & D. Zhang (Eds.), *Advances in Biometric Person Authentication* (Vol. 3781, pp. 135–141). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/11569947_17