

The Impact of Age-Related Variables on Facial Comparisons with Images of Children: Algorithm and Practitioner Performance

Dana Jaclyn Michalski

B.Psych (Hons) B.BehSc.

School of Psychology
University of Adelaide

November 2017



THE UNIVERSITY
of ADELAIDE

Thesis submitted for the degree of Doctor of Philosophy

Contents

Abstract	<i>i</i>
Declaration	<i>iii</i>
Acknowledgements	<i>iv</i>
Chapter 1. Research Overview	1
1.1 Introduction	1
1.2 Overview of the Thesis.....	3
1.3 Determining the Identity of Children in National Security Agencies.....	5
1.4 Facial Comparisons – Algorithms	6
1.4.1 Introduction to Biometric Systems.....	6
1.4.2 Facial Recognition Systems.....	8
1.5 Facial Comparisons – Practitioners.....	12
1.5.1 Face Recognition and Face Matching conducted by the General Population	12
1.5.2 Facial Comparison Practitioners	13
1.5.3 Limitations of Facial Comparison Research	15
1.6 Age Defined in the Context of this Thesis	17
1.7 Approaches to the Issue of Age Variation	18
1.8 Age-Related Facial Changes throughout the Lifespan.....	20
1.9 Thesis Aims	23
Chapter 2. Study 1: Requirements Collection from System Administrators, Facial Comparison Practitioners, and Algorithm Vendors	24

2.1	<i>Introduction</i>	24
2.2	<i>Method</i>	26
2.2.1	<i>Sampling</i>	26
2.2.2	<i>Design</i>	28
2.2.3	<i>Data Collection</i>	28
2.3	<i>Analysis</i>	33
2.4	<i>Findings</i>	34
2.4.1	<i>Focus Groups and Observation Sessions conducted at Agencies</i>	35
2.4.2	<i>Surveys Completed by Algorithm Vendors</i>	43
2.5	<i>Discussion</i>	47
2.6	<i>Summary</i>	51

Chapter 3. *A Methodological Primer: Image Preparation, Justification for the Methods Adopted, Performance Measures, and Analytical Techniques. 53*

3.1	<i>Introduction</i>	53
3.1.1	<i>A Controlled Operational Facial Image Database</i>	55
3.1.2	<i>Failure-to-Enrol Rate of the Database</i>	58
3.1.3	<i>Data Integrity Checking of the Database and Pre-Testing of the Algorithms</i>	58
3.1.4	<i>Secure Storage of the Database</i>	59
3.1.5	<i>Structure of the Database</i>	59
3.2	<i>Conduct of Algorithm Studies and Facial Comparison Practitioner Studies</i>	62
3.3	<i>Algorithm Studies</i>	63
3.3.1	<i>Selection of Images</i>	63
3.4	<i>Explanation of Variables used to Measure Algorithm Performance</i>	65
3.4.1	<i>False Match Rate (FMR)</i>	65
3.4.2	<i>False Non-Match Rate (FNMR)</i>	66
3.5	<i>Data Analysis Techniques for Algorithm Studies</i>	66
3.5.1	<i>Detection Error Trade-off (DET) Curves</i>	66
3.5.2	<i>Cumulative Probability Plots</i>	67
3.5.3	<i>Heat Map Data Matrices</i>	68
3.6	<i>Facial Comparison Practitioner Studies</i>	70
3.6.1	<i>Ratio of Mated to Non-Mated Image Pairs</i>	70

3.6.2	Pre-Selection of Image Pairs	71
3.7	<i>Explanation of Variables used to Measure Facial Comparison Practitioner</i>	
	<i>Performance</i>	75
3.7.1	Accuracy	76
3.7.2	Confidence.....	76
3.7.3	Response Time.....	77
3.7.4	Discrimination and Bias	77
3.8	<i>Data Analysis Techniques used to Measure Facial Comparison Practitioner</i>	
	<i>Performance</i>	78
3.8.1	Notched Boxplots	78
3.8.2	Heat Map Data Matrices	78
3.8.3	Statistical Analyses	79
3.9	<i>Summary</i>	79
Chapter 4.	<i>Study 2A: Facial Comparison Performance with Images of Children and Adults — Algorithm Study</i>	80
4.1	<i>Introduction</i>	80
4.2	<i>Research Questions</i>	83
4.3	<i>Methodology</i>	84
4.3.1	Design	84
4.3.2	Participants.....	84
4.3.3	Materials.....	85
4.3.4	Image Pair Selection	86
4.3.5	Procedure	87
4.4	<i>Results</i>	88
4.4.1	Algorithm Performance with Images of Children and Adults	88
4.4.2	Algorithm Performance with Images of Children and Adults on Mated and Non-Mated Image Pairs.....	90
4.4.3	Algorithm Performance with Images of Children and Adults based on set False Match Rates.....	92
4.5	<i>Discussion</i>	96
4.5.1	Algorithm Performance with Images of Children and Adults.....	96

4.5.2	Algorithm Performance with Images of Children and Adults on Mated and Non-Mated Image Pairs.....	97
4.5.3	Algorithm Performance with Images of Children and Adults based on set False Match Rates.....	98
4.5.4	Summary.....	99

Chapter 5. Study 2B: Facial Comparison Performance with Images of Children and Adults — Practitioner Study..... 100

5.1	Introduction	100
5.2	Research Questions.....	104
5.3	Methodology.....	105
5.3.1	Design	105
5.3.2	Participants.....	106
5.3.3	Materials.....	106
5.3.4	Image Pair Selection	109
5.3.5	Procedure	112
5.4	Results.....	114
5.4.1	Data Screening and Assumption Checking	115
5.4.2	Facial Comparison Practitioner Performance with Images of Children and Adults... ..	116
5.4.3	Facial Comparison Practitioner Performance with Images of Children and Adults on Mated and Non-Mated Image Pairs	120
5.5	Discussion	124
5.5.1	Facial Comparison Practitioner Performance with Images of Children and Adults... ..	124
5.5.2	Facial Comparison Practitioner Performance with Images of Children and Adults on Mated and Non-Mated Image Pairs	125
5.6	Summary.....	126

Chapter 6. Study 3A: Facial Comparison Performance with Images of Children at Different Ages and Age Variations — Algorithm Study 128

6.1	Introduction	128
6.2	Research Questions.....	131

6.3	<i>Methodology</i>	132
6.3.1	Participants.....	132
6.3.2	Materials.....	133
6.3.3	Image Pair Selection	133
6.3.4	Procedure	133
6.4	<i>Results</i>	133
6.4.1	Algorithm Performance with Images of Children at Different Ages and Age Variations.....	134
6.4.2	Algorithm Performance with Images of Children at Different Ages and Age Variations on Mated and Non-Mated Image Pairs	145
6.4.3	Algorithm Performance with Images of Children at Different Ages and Age Variations based on set False Match Rates	156
6.5	<i>Discussion</i>	160
6.5.1	Algorithm Performance with Images of Children at Different Ages and Age Variations.....	161
6.5.2	Algorithm Performance with Images of Children at Different Ages and Age Variations on Mated and Non-Mated Image Pairs	162
6.5.3	Algorithm Performance with Images of Children at Different Ages and Age Variations based on set False Match Rates	163
6.6	<i>Summary</i>	166
Chapter 7.	<i>Study 3B: Facial Comparison Performance with Images of Children at Different Ages and Age Variations — Practitioner Study</i>	167
7.1	<i>Introduction</i>	167
7.2	<i>Research Questions</i>	171
7.3	<i>Methodology</i>	172
7.3.1	Participants.....	172
7.3.2	Materials.....	173
7.3.3	Image Pair Selection	173
7.3.4	Procedure	177
7.4	<i>Results</i>	179
7.4.1	Practitioner Performance with Images of Children at each Age and Age Variation	179

7.4.2	Practitioner Performance with Images of Children at each Age and Age Variation on Mated and Non-Mated Image Pairs	187
7.4.3	Strategies Adopted by Practitioners to make Facial Comparison Decisions	194
7.5	<i>Discussion</i>	196
7.5.1	Practitioner Performance with Images of Children at each Age and Age Variation ..	198
7.5.2	Practitioner Performance with Images of Children for each Age and Age Variation on Mated and Non-Mated Image Pairs	200
7.5.3	Strategies Adopted by Practitioners to make Facial Comparison Decisions	204
7.5.4	Summary.....	206

Chapter 8. Study 4: Facial Comparison Performance for Agency Specific

Requirements — Practitioner and Algorithm Study..... 208

8.1	<i>Introduction</i>	208
8.1.1	Mock Example	209
8.2	<i>Research Questions</i>	212
8.3	<i>Methodology</i>	214
8.4	<i>Results</i>	214
8.4.1	Algorithm Performance by Age Group and Renewal Period	214
8.4.2	Algorithm Performance by Age Group and Renewal Period on Mated and Non-Mated Image Pairs.....	216
8.4.3	Algorithm Performance by Age Group and Renewal Period based on set False Match Rates.....	218
8.4.4	Practitioner Performance by Age Group and Renewal Period	220
8.4.5	Practitioner Performance by Age Group and Renewal Period on Mated and Non-Mated Image Pairs.....	222
8.5	<i>Discussion</i>	225
8.5.1	Algorithms	225
8.5.2	Practitioners	228
8.5.3	Summary.....	229

Chapter 9. Summary, Recommendations, and Conclusion 231

9.1	<i>Introduction</i>	231
-----	---------------------------	-----

9.2	<i>Key Findings: Facial Comparison Performance with Images of Children and Adults....</i>	234
9.3	<i>Key Findings: Facial Comparison Performance with Images of Children at Different Ages and Age Variations.....</i>	239
9.4	<i>Key Findings: Facial Comparison Performance for Agency Specific Requirements</i>	244
9.5	<i>Limitations of the Research</i>	246
9.6	<i>Recommendations</i>	248
9.7	<i>Contribution of the Research</i>	259
9.8	<i>Conclusion.....</i>	261
	References	263
	Appendices	287

List of Figures

Figure 1.	Overview of the thesis.	3
Figure 2.	Image of a child holding their current passport.	5
Figure 3.	Components of a generic biometric system (ISO/IEC 19795-1:2006, 2006).	8
Figure 4.	Age-related facial changes in siblings from birth to adulthood. Age of female sibling in top row: birth, 10 months, 2.5 years, 6 years, 10 years, 14 years, 18 years, and 25 years. Age of male sibling in bottom row: birth, 7 months, 2 years, 5 years, 9 years, 15 years, 19 years, and 25 years.	21
Figure 5.	Bone growth across childhood. Skull at age: 4 months (left), 1.5 years (middle), and 13 years (right).	21
Figure 6.	Example of the types of images that were and were not accepted into the database.	56
Figure 7.	The structure of the controlled facial image database for Study 2A and 2B.	60
Figure 8.	Example of a Detection Error Trade-off (DET) curve.	67
Figure 9.	Example of a cumulative probability plot.	68
Figure 10.	Example of a heat map data matrix of false non-match rates at a false match rate of 0.001.	69
Figure 11.	Screenshot of plotting program. Different variables (filters) can be applied to create the required plots.	86
Figure 12.	DETs displaying performance of each algorithm across Child and Adult groups.	89
Figure 13.	Cumulative probability plots displaying performance of each algorithm across Child and Adult groups.	91
Figure 14.	False non-match rates at false match rates of 0.1, 0.01, and 0.001 by Child and Adult groups.	93
Figure 15.	Screenshot of Comparer. Images are for illustration purposes only.	107

Figure 16.	<i>Image pairs presented to the facial comparison practitioners.....</i>	110
Figure 17.	<i>An example of the experimental screen layout for the facial comparison practitioner studies. Images are for illustration purposes only.....</i>	114
Figure 18.	<i>Overall accuracy and confidence for the Child and Adult groups.....</i>	117
Figure 19.	<i>Response times for the Child and Adult groups.....</i>	118
Figure 20.	<i>Accuracy for mated or non-mated Child and Adult groups.....</i>	120
Figure 21.	<i>Confidence for mated and non-mated Child and Adult groups.....</i>	121
Figure 22.	<i>Response times for mated and non-mated Child and Adult groups.....</i>	123
Figure 23.	<i>DETs for each algorithm based on age.....</i>	135
Figure 24.	<i>DETs for Algorithm E displaying how age impacts on performance for age variations spanning 0–5 years.....</i>	137
Figure 25.	<i>DETs for Algorithm E displaying how age impacts on performance for age variations spanning 6–10 years.....</i>	138
Figure 26.	<i>DETs for each algorithm based on age variation (AV = age variation in years). 140</i>	140
Figure 27.	<i>DETs for Algorithm E displaying how age variations impact on performance for ages 0–5 (AV = age variation in years).....</i>	142
Figure 28.	<i>DETs for Algorithm E displaying how age variations impact on performance for ages 6–11 (AV = age variation in years).....</i>	143
Figure 29.	<i>DETs for Algorithm E displaying how age variation impacts on performance for ages 12–17 (AV = age variation in years).....</i>	144
Figure 30.	<i>Cumulative probability plots for each algorithm based on age.....</i>	146
Figure 31.	<i>Cumulative probability plots for Algorithm E displaying how age impacts on performance for age variations spanning 0–5 years.....</i>	148
Figure 32.	<i>Cumulative probability plots for Algorithm E displaying how age impacts on performance for age variations spanning 6–10 years.....</i>	149
Figure 33.	<i>Cumulative probability plots for each algorithm based on age variation (AV = age variation in years).....</i>	151
Figure 34.	<i>Cumulative probability plots for Algorithm E displaying how age variation impacts on performance for ages 0–5 (AV = age variation in years).....</i>	153
Figure 35.	<i>Cumulative probability plots for Algorithm E displaying how age variation impacts on performance for ages 6–11 (AV = age variation in years).....</i>	154

Figure 36.	<i>Cumulative probability plots for Algorithm E displaying how age variation impacts on performance for ages 12–17 (AV = age variation in years).</i>	<i>155</i>
Figure 37.	<i>False match rate and false non-match rate data for Algorithm E based on a threshold set at a false match rate of 0.001 with images of adults.</i>	<i>157</i>
Figure 38.	<i>False non-match rate data for every age (0–17 years) and age variation (0–10 years) for Algorithm E based on a false match rate of 0.001 based on images of children at each of these ages and age variations.</i>	<i>159</i>
Figure 39.	<i>The 198 different categories in this study based on age and age variation with 120 image pairs selected per category. The circled cell indicates 120 image pairs contain a child at age 6 and an image of a child 4 years older.</i>	<i>174</i>
Figure 40.	<i>Example of the type of images used for the mated and non-mated pairs. The image of the child at the youngest age in the mated pair is also used as the youngest child in the non-mated pairs. Images are for illustration purposes only.</i>	<i>176</i>
Figure 41.	<i>Overall accuracy for each age and age variation (%).</i>	<i>180</i>
Figure 42.	<i>Overall confidence for each age and age variation (%).</i>	<i>182</i>
Figure 43.	<i>Overall response times for each age and age variation (seconds).</i>	<i>184</i>
Figure 44.	<i>Discrimination and bias for each age and age variation.</i>	<i>186</i>
Figure 45.	<i>Accuracy for mated and non-mated image pairs for each age and age variation (%).</i>	<i>188</i>
Figure 46.	<i>Confidence for mated and non-mated image pairs for each age and age variation (%).</i>	<i>191</i>
Figure 47.	<i>Response times for mated and non-mated image pairs for each age and age variation (seconds).</i>	<i>193</i>
Figure 48.	<i>DETs displaying performance of each algorithm across the six groups (A = age, RP = renewal period).</i>	<i>215</i>
Figure 49.	<i>Cumulative probability plots displaying performance of each algorithm across the six groups (A = age, RP = renewal period).</i>	<i>217</i>
Figure 50.	<i>False non-match rate for the six groups based on a false match rate of 0.001 (A = age, RP = renewal period).</i>	<i>219</i>
Figure 51.	<i>Overall practitioner accuracy for different groups (A= age, RP = renewal period).</i>	<i>221</i>
Figure 52.	<i>Practitioner accuracy (± 1 standard error) for different groups based on pair type (A = age, RP = renewal period).</i>	<i>222</i>

Figure 53. Accuracy and confidence based on pair type across age variation. 243

List of Tables

Table 1.	<i>Subsystems in a Biometric System and their Role (Martin, 2013).....</i>	<i>7</i>
Table 2.	<i>Categories of Facial Comparison Practitioners and their Role</i>	<i>13</i>
Table 3.	<i>Attendees at each Focus Group by Agency.....</i>	<i>30</i>
Table 4.	<i>Resources used by each Agency to make One-to-One Facial Comparisons</i>	<i>35</i>
Table 5.	<i>Ages Currently Examined by each Agency and the Ages Required to be Examined for Business Objectives</i>	<i>38</i>
Table 6.	<i>Criteria and Justification for Selection of Image Pairs.....</i>	<i>72</i>
Table 7.	<i>Variables used to Measure Facial Comparison Practitioner Performance</i>	<i>76</i>
Table 8.	<i>False Match Rate and False Non-Match Rate of the Child group when the Adult False Match Rate was set at 0.1, 0.01, and 0.001.....</i>	<i>95</i>
Table 9.	<i>Discrimination and Bias for the Child and Adult Groups</i>	<i>119</i>
Table 10.	<i>Strategies/Features Adopted to make Facial Comparison Decisions</i>	<i>195</i>
Table 11.	<i>Discrimination and Bias for Each Group.....</i>	<i>224</i>
Table 12.	<i>Algorithm Performance at a set False Match Rate of 0.001 based on Images of Adults or Children</i>	<i>235</i>
Table 13.	<i>Practitioners Performance for Child and Adult groups based on Mated and Non-Mated Image Pairs.....</i>	<i>238</i>
Table 14.	<i>Algorithm C's False Non-Match Rate Performance when the False Match Rate is set at 0.001 based on Different Image Groups.....</i>	<i>240</i>
Table 15.	<i>Practitioner Accuracy (%) by Age (Years)</i>	<i>241</i>
Table 16.	<i>Practitioner Accuracy (%) across Age (Years) based on Pair Type.....</i>	<i>242</i>

Abstract

Determining the identity of children is critical for many national security agencies for example, to aid in the fight against child exploitation, trafficking, and radicalised minors, as well as for passport control and visa issuance purposes. Facial comparison is one method that may be used to achieve this. Facial comparison can be conducted using an algorithm (within a facial recognition system), manually by a facial comparison practitioner, or by a combination of the two. Much of the previous research examining facial comparison performance of both algorithms and practitioners has been conducted using images of adults. Due to the substantial amount of age-related facial growth that occurs in childhood, compared to adulthood, it is likely that performance will be poorer with images of children. The overarching aim of the research therefore, was to determine the impact of age-related variables, namely chronological age and age variation (the age difference between images) on facial comparison performance of algorithms and practitioners with images of children.

Study 1 involved consultation with national security agencies and algorithm vendors to identify the key requirements to examine in this thesis. After reviewing the literature to identify research gaps, five empirical studies were conducted. To ensure the studies were as operationally relevant as possible, a large database containing several million controlled images of children and adults was sourced, and five state-of-the-art facial recognition algorithms were employed. In addition, facial comparison practitioners from a government agency participated in the practitioner studies. Study 2A compared algorithm performance with images of children to performance with images of adults. Study 2B compared practitioner performance with images of children to performance with images of adults. Study 3A examined algorithm performance with images of children at each chronological age in childhood (0–17 years) and age variations ranging from 0–10 years apart. Study 3B examined

practitioner performance on the same age-related variables examined in Study 3A. Study 4 demonstrated how the data collected in Study 3A and 3B could be used to answer agency specific questions.

This thesis concludes with a series of recommendations for both the algorithm and practitioner domains, as well as future research directions designed to improve knowledge and performance regarding facial comparisons with images of children.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

In addition, I certify that no part of this work will, in the future, be used in a submission for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968. I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

.....

Dana Michalski

November 2017

Acknowledgements

I would like to thank my PhD supervisors, Dr Carolyn Semmler and Dr Rebecca Heyer for supporting me during my candidature. Dr Carolyn Semmler, thank you for agreeing to take me on and your words of encouragement over the last few years. Your enthusiasm and advice has made this experience more rewarding. Dr Rebecca Heyer, thank you for your support as a PhD and work-based supervisor. Your work ethic and passion to research keeps me motivated when things get tough. I hope we can continue to work together well into the future.

This PhD would not have been possible without the encouragement of Dr Brett McLindin. Without your ability to convince me that “doing a PhD is easy” (ha!), I would never have considered undertaking a PhD that crosses two distinct disciplines. Once again, you threw me in the deep end and because of that, I have become a stronger researcher.

I would also like to acknowledge the Defence Science and Technology Group (DST Group) and members of the Biometrics Team. Thank you to all team members who have provided support along the journey. A special mention must go to John Stwein and Martyn Hole for providing assistance through the development of a special plotting program that could handle the large amounts of algorithm data thrown at it. The constant back and forth to get the plotting program developed just right seemed like an additional burden at a time when there was already chaos. The final product is something you should both be proud of and will be used in future research by the team. To Peter Aquilina and Dr Kathy Hanton, thank you for taking the time to view the 23,760 image pairs selected for Study 3B. Although not a fun task, this helped to ensure rigour in the dataset selected. Sau Yee Yiu and Dr Veneta MacLeod, thank you for taking the time to read my thesis during the final draft. Your editorial comments helped to make this a more succinct thesis. A special mention also goes to Chris Malec. Thank you for

your support as I grappled with learning new terminology, methodology, and analyses techniques used in the biometrics world that completely differs to that from the psychology domain. Also, for your support throughout my PhD by persistently asking “are you broken yet?” I hope to return the favour to you one day soon! To Dr Robin Nicholson, thank you for supporting my research and providing me the opportunity to present at Biometrics 2016 in London (even if it felt like you didn’t have a choice).

The operational validity of this research was made possible by the generosity of agencies and industry. Thank you to each Australian and international agency who took the time to meet with me and provide requirements for this research. Your insights have made this an area that I have become passionate about. The studies would not have been possible on such a large scale without access to an already established database. Thank you to the providers of this database as it has allowed a depth of knowledge not possible before. To each of the practitioners who participated, thank you for taking the time to make this research possible. I hope the research will be used to make your jobs easier in the future. To the Westbourne Group, thank you for providing assistance in the development of an experimental application for the practitioner studies. Also, thank you to the algorithm vendors: 3M, Aware, Cognitec, Morpho, and NEC who provided their facial recognition algorithms for research purposes under Material Transfer Agreements. Your continued support helps agencies and researchers better understand the current state of facial recognition.

To all my family and close friends, thank you for your support and understanding along the journey. Remember all those times you asked what I was doing? Well now you can just read the next 300+ pages to find out! Finally, a special thank you goes to Jaimee Spurr. Without your encouragement, compassion, and food I would not have made it.