

https://dspace.org/

Reliability of forensic facial identification from CCTV recordings with implications for admissibility to court

Bacci, Nicholas

2021

https://hdl.handle.net/10539/32060

Downloaded from a DSpace repository. https://dspace.org/

Reliability of forensic facial identification from CCTV recordings with implications for admissibility to court



UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

Nicholas Bacci

472770

Original published work submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Doctor of Philosophy.

Supervisors:

Dr Nanette Briers, Senior Lecturer, School of Anatomical Sciences.

Prof Maryna Steyn, Professor, School of Anatomical Sciences.

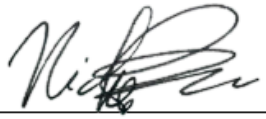
FACULTY OF HEALTH SCIENCES

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

18 JUNE 2021

Declaration: Student's contribution to article(s) and agreement of co-author(s)

I, Nicholas Bacci, student number 472770, declare that this Thesis is my own work and that I contributed adequately towards research findings published in the article(s) stated below which are included in my Thesis.

Signature of Student  **Date** 10/06/2021

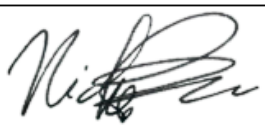
Name of Primary Supervisor: Nanette Briers

Signature of Primary Supervisor  **Date** 10/06/2021

Agreement by co-authors: By signing this declaration, the co-authors listed below agree to the use of the article(s) by the student as part of his Thesis. In cases where the student is not the 1st author of a published article, the primary supervisor must explain (under comments) why the student is entitled to use the paper for his/her degree purposes.

Article 1:
Title: Development of the Wits Face Database: an African database of high-resolution facial photographs and multimodal closed-circuit television (CCTV) recordings

Journal name, year, volume and page numbers:
F1000 Research, 2021. https://doi.org/10.12688/f1000research.50887.1

Authors	Name	Signature	Date
1st author	Nicholas Bacci		10/06/2021
2nd author	Joshua G Davimes		10/06/2021
3rd author	Maryna Steyn		10/06/2021
4th author	Nanette Briers		10/06/2021

Comments by primary supervisor:

The contribution distribution in this article was as follows:
Conceptualization: Nicholas Bacci, Maryna Steyn, Nanette Briers;
Data curation: Nicholas Bacci, Joshua Davimes;
Methodology: Nicholas Bacci, Joshua Davimes, Maryna Steyn, Nanette Briers;
Investigation: Nicholas Bacci, Joshua Davimes;
Project Administration: Nicholas Bacci, Joshua Davimes;
Resources: Nicholas Bacci, Joshua Davimes;
Visualisation: Nicholas Bacci, Joshua Davimes;
Writing - original draft preparation: Nicholas Bacci;
Writing - review and editing: Nicholas Bacci, Joshua Davimes, Maryna Steyn, Nanette Briers;
Funding acquisition: Nicholas Bacci, Maryna Steyn, Nanette Briers;
Supervision: Maryna Steyn and Nanette Briers.

Article 2:

Title: Validation of forensic facial comparison by morphological analysis in photographic and CCTV samples

Journal name, year, volume and page numbers:
International Journal of Legal Medicine, 2021. https://doi.org/10.1007/s00414-021-02512-3

Authors	Name	Signature	Date
1st author	Nicholas Bacci		10/06/2021
2nd author	Tobias MR Houlton		10/06/2021
3rd author	Nanette Briers		10/06/2021
4th author	Maryna Steyn		10/06/2021

Comments by primary supervisor:

The contribution distribution in this article was as follows:
Conceptualization: Nicholas Bacci, Maryna Steyn, Nanette Briers;
Methodology: Nicholas Bacci, Maryna Steyn, Nanette Briers, Tobias Houlton;
Formal analysis and investigation: Nicholas Bacci, Tobias Houlton;
Writing - original draft preparation: Nicholas Bacci;
Writing - review and editing: Nicholas Bacci, Maryna Steyn, Nanette Briers, Tobias Houlton;
Funding acquisition: Nicholas Bacci, Maryna Steyn, Nanette Briers;
Resources: Maryna Steyn, Nanette Briers; Supervision: Maryna Steyn, Nanette Briers.

Article 3:

Title: Performance of forensic facial comparison by morphological analysis across optimal and suboptimal CCTV settings

Journal name, year, volume and page numbers:

Science & Justice, 2021, submitted - under review

Authors	Name	Signature	Date
1st author	Nicholas Bacci		10/06/2021
2nd author	Maryna Steyn		10/06/2021
3rd author	Nanette Briers		10/06/2021

Comments by primary supervisor:

The contribution distribution in this article was as follows:
Conceptualization: Nicholas Bacci, Maryna Steyn, Nanette Briers;
Methodology: Nicholas Bacci, Maryna Steyn, Nanette Briers;
Formal analysis and investigation: Nicholas Bacci;
Writing - original draft: Nicholas Bacci;
Writing - review and editing: Nicholas Bacci, Maryna Steyn, Nanette Briers;
Funding acquisition: Nicholas Bacci, Maryna Steyn, Nanette Briers;
Resources: Maryna Steyn, Nanette Briers;
Supervision: Maryna Steyn, Nanette Briers.

Article 4:

Title: Assessing the effect of facial disguises on forensic facial comparison by morphological analysis

Journal name, year, volume and page numbers:

Journal of Forensic Sciences, 2021, https://doi.org/10.1111/1556-4029.14722.

Authors	Name	Signature	Date
1st author	Nicholas Bacci		10/06/2021
2nd author	Nanette Briers		10/06/2021
3rd author	Maryna Steyn		10/06/2021

Comments by primary supervisor:

The contribution distribution in this article was as follows:
Conceptualization: Nicholas Bacci, Maryna Steyn, Nanette Briers;
Methodology: Nicholas Bacci, Maryna Steyn, Nanette Briers;
Formal analysis and investigation: Nicholas Bacci;
Writing - original draft: Nicholas Bacci;
Writing - review and editing: Nicholas Bacci, Maryna Steyn, Nanette Briers;
Funding acquisition: Nicholas Bacci, Maryna Steyn, Nanette Briers;
Resources: Maryna Steyn, Nanette Briers;
Supervision: Maryna Steyn, Nanette Briers.

To my family, that has supported and encouraged me through my academic strides

Marianna Morfi

Corinna Bacci

Brando Bacci

Στην πολυαγαπημένη μου γιαγιά , που ίσως δεν καταλαβαίνει τα Αγγλικά ή αυτή τη διατριβή, αλλά ήταν πάντα περήφανη για τα επιστημονικά μου ενδιαφέροντα και ήλπιζε να με δει ερευνητή από τότε που ήμουν παιδί

Ελβίρα Μόρφη

Presentations arising from this project

1. Bacci N, Steyn M, Briers N. 2021 *Reliability of facial identification methods in optimal photographs and suboptimal CCTV footage.* American Association of Physical Anthropologists 90th Annual Meeting, April 15, 2021, virtual online conference.

The above presentation was originally submitted at the 2019 International Association for Craniofacial Identification Biennial Meeting and accepted for a podium presentation; however, the conference was cancelled due to hurricane Barry making landfall on the dates of the conference.

The above presentation was also submitted to the 2020 American Association of Physical Anthropologists 89th Meeting and accepted for a podium presentation. This meeting was, however, also cancelled, due to the COVID-19 pandemic surge through the world in April of 2020.

Publications arising from this project

1. Bacci N, Davimes J, Steyn M and Briers N. Development of the Wits Face Database: an African database of high-resolution facial photographs and multimodal closed-circuit television (CCTV) recordings [version 1; peer review: 1 approved]. F1000Research 2021, 10:131. https://doi.org/10.12688/f1000research.50887.1

2. Bacci N, Houlton TMR, Briers N, Steyn M. Validation of forensic facial comparison by morphological analysis in photographic and CCTV samples. Int J Legal Med 2021. https://doi.org/10.1007/s00414-021-02512-3

3. Bacci N, Briers N, Steyn M. Assessing the effect of facial disguises on forensic facial comparison by morphological analysis. J Forensic Sci 2021. https://doi.org/10.1111/1556-4029.14722

Abstract

Forensic facial comparison aims to associate a suspect to a crime scene through photographic or surveillance data. Morphological analysis (MA) of facial features is considered the go-to method for forensic facial comparison (FFC) as other techniques underperform in uncontrolled real-life scenarios. Current MA guidelines include a yet to be rigorously tested facial feature list developed by the Facial Identification Scientific Working Group (FISWG). The present study aimed at validating facial comparison by MA with the use of the FISWG facial component list.

The sample comprised of a multimodal high-resolution photographic and video recording face database specifically developed for this study. Standardised and matching non-standardised facial photographs (n=6220) and CCTV video recordings (n=334) of 622 consenting informed participants were compiled into the Wits Face Database. CCTV recordings included standard CCTV setups with an average surveillance camera, high-definition eye-level height CCTV camera and a low-resolution analogue camera. Some participants were also recorded wearing sunglasses or caps. Photographs and images from the recordings were compiled into 485 face pools. Each face pool contained one target image contrasted to 10 potential matching images. Nine of the potential matches were foils and one was a correct match across most face pools, approximately 9% contained no correct matches to prevent forced matches. Target images were individually compared to each potential match using the FISWG feature list and the South African Police Services criteria. Trial outcomes were analysed in confusion matrices and test performance statistics were computed.

MA was found to be highly accurate and reliable for FFC under optimal conditions such as photographs, high-resolution eye-level CCTV, and standard CCTV, while underperforming across analogue CCTV and disguised faces. False positives were less common than false negatives. These results agreed with other studies identifying feature lists to benefit the facial comparison process. FFC in sub-optimal settings was less accurate. Facial information was lost when recording conditions were subpar and when obstructed by caps with drastically decreased accuracy and increased false negatives.

This study has established MA and the FISWG feature list as a valid and accurate method for use in forensic facial identification when high quality data are available.

Acknowledgements

This PhD was conducted with support from the South African National Research Foundation (NRF) (DAAD-NRF Scholarship, Grant No.: 11858) and the J.J.J. Smieszek Fellowship from the School of Anatomical Sciences, University of the Witwatersrand awarded to N. Bacci as well as NRF funds awarded to N. Briers as part of the Improving Methodologies and Practices in Craniofacial Identification (Grant No.: CSUR160425163022; UID: 106031).

I humbly thank all the participants who agreed to be photographed and recorded for the development of the Wits Face Database and from which the facial comparison analyses were conducted. Thanks are due to Dr Joshua Davimes for his instrumental photography and project management aid in the development of the Wits Face Database, Mr Gideon LeRoux for his assistance in the capturing and extracting of CCTV recordings from the University security systems. Thanks are owed to Ms Tamara Lottering for her assistance in composing the face pools and Dr Tobias Houlton for aiding in study validation and facial comparison training. Thanks are also due to the volunteers who assisted in participant recruitment: Jesse Fredericks, Kiveshen Pillay, Rethabile Masiu, Sameerah Sallie, Daniel Munesamy, Laurette Joubert, Jordan Swiegers, Betty Mkabela, Johannes P. Meyer, Amy Spies, Natasha Loubser, Nicole Virgili, Dan-Joel Lukumbi, Tamara Lottering, Mathabatha Ntjie, Claudia Landsman, Raheema Dalika, Merete Goosen, Stephanie Souris, Rabelani Negota, Mahlatse Mahasha, Jessica Manavhela.

I direct many thanks to my supervisors, Dr Nanette Briers and Prof Maryna Steyn for their backing, expertise, guidance, reliability, and support throughout this challenging endeavour from project inception to thesis submission! They provided me many opportunities to grow and learn in forensics and science in general. While teaching me how to identify faces and human remains, I also learnt a lot about identifying myself.

I greatly thank Dr Tanya Augustine and Dr Brendon Billings for their friendship, mentoring, guidance, and emotional support in navigating and coping with the difficulties of doing a PhD and being a budding academic. Their influence has truly made me a better person and hopefully a successful and productive academic in future.

A great deal of thanks is owed to my wife and best friend, Pratheeksha Joseph-Bacci, for always being immeasurably patient, encouraging, and supportive in every capacity, always and forever.

Table of Contents

Declaration i

Dedication v

Presentations arising from this project vi

Publications arising from this project vii

Abstract viii

Acknowledgements ix

List of Abbreviations xii

List of Figures xiii

List of Tables xv

1. Introduction – Integrated Narrative 1

1.1 Photography and crime 1

1.2 From photographs to surveillance videos 2

1.3 Facial identification from digital evidence 2

1.4 Automated facial recognition 4

1,5 Forensic facial comparison 4

1.6 Approaches to forensic facial comparison 5

1.7 Morphological analysis as best practice 9

1.8 Limitations of CCTV in facial identification 12

1.9 Image quality 12

1.10 Disguises and facial identification 14

1.11 The need for facial image databases 14

1.12 Aims and objectives 15

References – Integrated narrative 17

2. Development of the Wits Face Database: an African database of high-resolution facial photographs and multimodal closed-circuit television (CCTV) recordings 29

3. Validation of forensic facial comparison by morphological analysis in photographic and CCTV samples 48

4. Performance of forensic facial comparison by morphological analysis across optimal and suboptimal CCTV settings 84

5. Assessing the effect of facial disguises on forensic facial comparison by morphological analysis 119

6. Concluding discussion 154

6.1 Morphological analysis validation 155

6.2 Facial comparison across CCTV conditions 156

6.3 Effects of disguises on facial comparison 158

6.4 Limitations 160

6.5 Future directions 163

6.6 Concluding remarks 165

References - Concluding discussion 167

Appendix A – Ethics Clearance Certificate 172

Appendix B -Tabulated Facial Comparison Feature List (FISWG, 2018) 173

Appendix C – Supplementary Material Manuscript 1 184

Appendix D - Supplementary Material Manuscript 2 198

Appendix E - Supplementary Material Manuscript 3 204

Appendix F - Supplementary Material Manuscript 4 205

Appendix F - Publication Licences & Copyright 206

Appendix G - Plagiarism Declaration 215

Appendix H – Turnitin Report 216

List of Abbreviations

Abbreviation	Full name
CCTV	Closed-circuit television
FISWG	Facial Identification Scientific Working Group
FN	False Negative
FNR	False Negative Rate
FP	False Positive
FPR	False Positive Rate
IA	Inconclusive Analysis
MA	Morphological analysis
NIR	No Information Rate
NPV	Negative Predictive value
PPV	Positive Predictive Value
TN	True Negative
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate
WFD	Wits Face Database

List of Figures

Manuscript 1: Development of the Wits Face Database: an African database of high-resolution facial photographs and multimodal closed-circuit television (CCTV) recordings

Figure number	Figure title	Page
1	Schematic diagram of camera set-up for closed- circuit television (CCTV) and facial photograph capture.	35
2	Actual photographic camera set-up in the process of database development.	36
3	Example of the five views of standardised (left) and wildtype (right) facial photographs captured.	37
4	Demonstration of estimated objective to face distance at site A.	38
5	Example of the five views of standard closed-circuit television (CCTV) stills captured at site A	38
6	Example of the five views of eye level closed-circuit television (CCTV) stills captured at site B.	39
7	Example of the five views of analogue closed-circuit television (CCTV) stills captured at site C. The	39
8	Example of the five views of closed-circuit television (CCTV) recordings with two obstruction types - brimmed cap (left) and sunglasses (right).	40

Manuscript 2: Validation of forensic facial comparison by morphological analysis in photographic and CCTV samples

Figure number	Figure title	Page
1	Facial comparison analysis example	81
2	Bar graphs of the facial comparison outcomes under a lenient (a) and strict (b) iteration within the photograph-to-photograph cohort by trial	82

3	Bar graphs of the facial comparison outcomes under a lenient (a) and strict (b) iteration within the CCTV-to-photograph cohort by trial	83

Manuscript 3: Performance of forensic facial comparison by morphological analysis across optimal and suboptimal CCTV settings

Figure number	Figure title	Page
1	Facial comparison analysis example from the Digital CCTV sample	109
2	Facial comparison analysis example from the analogue CCTV sample	110
3	Bar graphs of the facial comparison outcomes under a lenient (A) and strict (B) iteration within the Digital CCTV to photograph cohort by trial	111
4	Bar graphs of the facial comparison outcomes under a lenient (A) and strict (B) iteration within the Analogue CCTV to photograph cohort by trial	112

Manuscript 4: Assessing the effect of facial disguises on forensic facial comparison by morphological analysis

Figure number	Figure title	Page
1	Facial comparison analysis example from the cap obstruction CCTV sample	150
2	Facial comparison analysis example from the glasses obstruction CCTV sample	151
3	Bar graphs of the facial comparison outcomes under a lenient (A) and strict (B) iteration within the cap CCTV cohort by trial	152
4	Bar graphs of the facial comparison outcomes under a lenient (A) and strict (B) iteration within the sunglasses CCTV cohort by trial	153

List of Tables

Manuscript 1: Development of the Wits Face Database: an African database of high-resolution facial photographs and multimodal closed-circuit television (CCTV) recordings

Table number	Table title	Page
1	Overview of available face database with available descriptives	33 – 34
2	Detailed categorisation of Wits Face Database (WFD) composition by cohorts of potential analyses	41

Manuscript 2: Validation of forensic facial comparison by morphological analysis in photographic and CCTV samples

Table number	Table title	Page
1	Facial pose distribution throughout face pool sample	76
2	Face pool trials by analysis: cohort subsample breakdown	72 - 73
3	Summarised list of facial components for facial comparison by morphological analysis [29]	77
4	Contingency table reference for confusion matrix of positive and negative matches (adapted from Kuhn [35]	77 - 78
5	Summary of intra- and inter-observer agreements across all cohorts of analysis	78
6	Contingency table for confusion matrix of photograph- to-photograph comparison lenient iteration	78
7	Contingency table reference for confusion matrix of CCTV-to-photograph comparison lenient iteration	78
8	Confusion matrix function results and chance- corrected accuracies across the lenient iteration of analysis	79
9	Tabulation of all predictive quality values calculated across the lenient iteration of analyses	79

10	Contingency table reference for confusion matrix of the strict iteration of the photograph-to-photograph comparisons analysed (left), compared to the ideal perfect matches and exclusions for direct comparison of performance of the facial comparisons (right)	79 - 80
11	Contingency table reference for confusion matrix of the strict iteration of the CCTV-to-photograph comparisons analysed (left), compared to the ideal perfect matches and exclusions for direct comparison of performance of the facial comparisons (right)	80
12	Confusion matrix function results across the strict iteration of each cohort of analysis	80
13	Tabulation of the predictive quality values calculated across the strict iteration of each cohort of analysis	80-81
14	False-positive rates (FPR) and false-negative rates (FNR) calculated across both lenient and the strict iterations of the analysis cohorts	81

Manuscript 3: Performance of forensic facial comparison by morphological analysis across optimal and suboptimal CCTV settings

Table number	Table title	Page
1	Face pool trials by analysis: cohort subsample breakdown.	104
2	Facial pose distribution throughout face pool sample.	104
3	Contingency table reference for confusion matrix of positive and negative matches (adapted from Kuhn [42])	105
4	Equation list for statistical analysis based on the Caret R package [41,42]	105
5	Tabulation of intra- and inter-observer agreements across cohorts of analysis.	105 - 106
6	Contingency table reference for confusion matrix of the lenient iteration of the Digital CCTV cohort	106

7	Contingency table reference for confusion matrix of the lenient iteration of the Analogue CCTV cohort analysed	106
8	Confusion matrix function results and chance corrected accuracies across cohorts of analysis	106 – 107
9	Tabulation of all predictive quality values calculated across cohorts of analyses	107
10	Contingency table reference for confusion matrix of the strict iteration of the Digital CCTV cohort	107
11	Contingency table reference for confusion matrix of the strict iteration of the Analogue CCTV cohort analysed	107 - 108
12	Confusion matrix function results across the strict iteration of each cohort of analysis	108
13	Tabulation of the predictive quality values calculated across the strict iteration of each cohort of analysis	108

Manuscript 4: Assessing the effect of facial disguises on forensic facial comparison by morphological analysis

Table number	Table title	Page
1	Analysis trials face pools: cohort subsample breakdown	144
2	Facial pose distribution throughout face pool sample	144
3	FISWG facial component list for facial comparison by morphological analysis (37)	144 - 145
4	Contingency table reference for confusion matrix of positive and negative matches adapted from Kuhn (46)	145
5	Tabulation of statistical metrics used based on the Caret R package (46,47), equations are available as supplementary information.	145 - 146
6	Contingency table for confusion matrix of the lenient iteration of the Cap CCTV cohort	146 – 147
7	Contingency table reference for confusion matrix of the lenient iteration of the Sunglasses CCTV cohort analysed	147

8	Confusion matrix function results and chance corrected accuracies across cohorts of analysis	147
9	Tabulation of the lenient iteration predictive quality values calculated across cohorts of analyses	147 – 148
10	Contingency table reference for confusion matrix of the strict iteration of the Cap CCTV cohort	148
11	Contingency table reference for confusion matrix of the strict iteration of the Sunglasses CCTV cohort analysed	148 - 149
12	Confusion matrix function results across the strict iteration of each cohort of analysis	149
13	Tabulation of predictive quality values calculated across the strict iteration of each cohort of analysis	149

1. Introduction – Integrated Narrative

South Africa ranks as the country with the 3rd highest crime index in the world and the highest in Africa (Adamovic, 2021). This is evidenced by the almost two million crimes reported between April 2019 and March 2020 alone, almost 1.6 million of which are considered part of the 17 serious crimes, including murder, assault, sexual assault, etc. (South African Police Services, 2020).

1.1 Photography and crime

A technological advancement that has historically aided in crime investigation, resolution and even prevention is the advent of cameras. Photography has played a role in surveillance, identification, and detection of criminals since the mid to late 19th century (Jäger, 2001). These applications were originally employed in various European cities by implementing anthropological standards to depict portraits of habitual criminals, in a modern day mugshot format, for police controlled registries (Jäger, 2001). The aim of these registers was to permit witnesses and victims to identify potential suspects by simulating the facial review of a face-to-face meeting (Jäger, 2001). However, this proved ineffective for identification and detections due to a lack of standardisation throughout the image capture process employed by the various individual police departments (Jäger, 2001).

In the 1890s, Bertillon attempted to better catalogue and identify individuals with a more methodical approach to "judicial photography" incorporating anthropometry, which led to standardising image capture conditions (Bertillon and McClaughry, 1896; Jäger, 2001). This image capture standardisation was achieved by photographing apprehended criminals in the now traditionally perceived mugshots, with anterior and lateral facial views with a neutral expression and stance (Bertillon and McClaughry, 1896; Jäger, 2001). This shift in image capture procedure allowed for an effective, scientific, and standardised method of photographing offenders. Standardised images were then used to aid in identification and detection of suspects in areas away from where the original crime took place. As the application and public knowledge of this "police photography" increased, it resulted in reduced crime rates (Jäger, 2001). While Bertillon's approach to facial anthropometry was soon replaced by fingerprint identification due to its many shortcomings and limitations (Faigman, 2008), the image capture standards persisted through the 1900s in many countries (Jäger, 2001).

A different application to photography was employed in the 20th century, where photographs were captured to facilitate crime scene investigations directly (Mokwena, 2012). Standardised photography permitted accurate details to be captured and demonstrated in a judicial context (Mokwena, 2012). The practice of including crime scene photographs in court proceedings and newspaper publications became more common between the 1930s and 1950s in both the United States of America and the United Kingdom (Bell, 2018).

1.2 From photographs to surveillance videos

Photography has been widely implemented for both identification and investigation purposes across the world. However, the technology that truly revolutionised legal investigations and facilitated progress in criminology as a whole, was closed-circuit television (CCTV) surveillance (Lindegaard and Bernasco, 2018). CCTV cameras are, mostly, perceived as an improvement of photography as they have advanced both identification and investigation in multiple ways. Since the 1990s, CCTV system installations have systematically increased in many countries, including South Africa (Norris *et al.*, 2002; Jain, *et al.*, 2012; Moyo, 2019; Piza *et al.*, 2019), to the extent where for many people they are considered the public norm (Goold *et al*., 2013; Duncan, 2018).

The main strategy behind CCTV implementation has been to decrease the opportunities of committing a crime and increase a criminal's perceived risks, such as by identification (Piza *et al.*, 2019). Although there is not always consensus on the exact local effects (Moyo, 2019), CCTV is thought to lead to localised crime reduction (Gill and Spriggs, 2005; Welsh and Farrington, 2008; Piza *et al.*, 2019). CCTV has also been demonstrated to be frequently useful in police investigations after crimes are committed, by acting as additional evidence for identification and monitoring the events of a crime scene (Ashby, 2017; Piza *et al.*, 2019). In fact, there has been clear indication that when CCTV data is available it is substantially more likely for criminal activity to be resolved (Ashby, 2017).

1.3 Facial identification from digital evidence

With the high levels of crime and widespread use and effectiveness of CCTV, the opportunity and need for facial identification has become more common worldwide (Kleinberg and Siebert, 2012) and is particularly in demand in South Africa (Steyn *et al.*, 2018). This increasing demand originates mostly from the large scale availability of image data, not only in the form

of CCTV (Norris *et al.*, 2002; Kleinberg and Siebert, 2012), but even as photographic and video recordings from cell phones (Houlton and Steyn, 2018).

Forensic facial identification is a form of facial imaging, a group of disciplines in forensic anthropology making use of facial visual data to aid in identification (Stephan *et al.*, 2019), routinely employed to associate suspects to criminal activities based on the analysis of photographic or video evidence (Steyn *et al.*, 2018). These disciplines include facial approximation, photographic superimposition, age progression and regression, molecular photofitting, facial composites and sketches, facial depiction, and facial comparison (Stephan *et al.*, 2019). Each of these methods serve specific purposes, with niche applications in various types of criminal circumstances. Some aim to generate additional facial image data from the limited information available, such as from genetic information from a scene (molecular photofitting), witness descriptions (facial composites and sketches), or craniometric data (facial approximation) (Stephan *et al.*, 2019). Other practices aim to be used for preliminary identification purposes from either crania (photographic superimposition) or actual facial images (forensic facial comparison) (Stephan *et al.*, 2019). While many of these disciplines have had extensive research conducted on their efficacy and improvement (e.g. facial approximation, facial composites and sketches) (Stephan *et al.*, 2019), facial comparison for identification remains largely untested despite its increasing demand (Steyn *et al.*, 2018).

Facial identification and facial recognition are two related yet distinct processes. Facial recognition is largely considered the intrinsic cognitive capacity of humans to recognise faces and is prone to high error rates (Bruce *et al.*, 2001; Boyce, Lindsay and Brimacombe, 2008), likely due to its holistic, rapid and unsystematic process (Schüler and Obertová, 2020). Facial identification, however, is the process o providing specific and virtually perfect correspondence of identity (Schüler and Obertová, 2020). While identification often arises and relies on human recognition ability, it requires further practical analysis involving standardised, detailed, comprehensive and meticulously recorded methodology (Schüler and Obertová, 2020). Facial identification is usually employed through the process of facial comparison. Facial comparison involves the “manual” human-based review of facial images for identity confirmation (Davis *et al.*, 2012; Valentine and Davis, 2015; Facial Identification Scientific Working Group, 2019).

1.4 Automated facial recognition

The manual facial comparison process is distinct from automated or semi-automated facial recognition, which employs various computer-based methods to attempt confirming facial identity (Kaur *et al.*, 2020). These systems are currently still reliant on human validation in their operating loop (Spaun, 2009). While the accuracy of automated facial recognition systems has seen an increase in recent years (Grother *et al.*, 2018), there are many concerns, particularly relating to high false positive matches (Dodd, 2018; Grother *et al.*, 2018, 2019), that need addressing before these systems can be employed at a wide scale in a legal context. A large number of these concerns arise from the methodological approaches of these systems relying on biometric information (White *et al.*, 2015) and highly standardised images (Davis *et al.*, 2010; Urbanová, 2016; Press Association, 2018), which are generally not available in the unstandardised set-up of most surveillance systems. In addition, due to a lack of varied and multimodal databases for testing these automated facial recognition algorithms, many of the available software have intrinsic biases, including population biases (Grother *et al.*, 2019; Wall, 2019), that further hinder their deployment in a realistic scenario. As a result, human observer-based facial image comparison is considered the go-to approach (Wilkinson and Evans, 2009; Davis *et al.*, 2012; Valentine and Davis, 2015; White *et al.*, 2015) and will likely still be considered the go-to validation method following improvement and widespread deployment of automated systems.

1,5 Forensic facial comparison

Facial comparison is involved in identification and tracking of unfamiliar faces and has numerous potential applications and approaches to it. The four primary applications include intelligence gathering irrespective of identity for management, access control and screening, tracking investigative information not intended to be presented in a judicial context, and forensic comparisons to be used in judicial proceedings (Facial Identification Scientific Working Group, 2019). The approach to each of these applications is highly dependent on the specific acceptable compromises and requirements of each circumstance. For example, in the well-known scenario of customs or passport control checkpoints, the high influx of people limits the available time permissible for a detailed facial comparison. As such, a rapid image-to-image or image-to-person assessment would then be carried out (Facial Identification Scientific Working Group, 2012, 2019), relying primarily on the innate human ability to perform a holistic comparison of faces. Additional approaches include facial review and facial

examination, both of which are more rigorous than a rapid facial assessment (Facial Identification Scientific Working Group, 2019). Facial review is based on image to image comparisons and is primarily employed in investigative and intelligence gathering operations with a wide range of thoroughness and specific applications, while facial examinations are intended to be strict, methodical image-to-image based comparisons that can be applied in a forensic setting (Facial Identification Scientific Working Group, 2019).

Both review and examination are recommended to include a minimum of one additional analyst verify the outcomes of the primary analysis (Facial Identification Scientific Working Group, 2019). Facial examination or forensic facial comparison should follow the Analysis, Comparison, Evaluation, and Verification (ACE-V) approach (Facial Identification Scientific Working Group, 2019), used in other forensic practices such a fingerprint identification (Speckeis, 2011). The ACE-V methodological approach is aimed at integrating principles of the scientific method in the act of performing any forensic comparison to augment its performance and reliability (Speckeis, 2011).

While the ACE-V method is considered crucial to comparisons, expertise in the specific field of forensic comparison at hand should not be discounted (Stevenage and Pitfield, 2016). This is of particular relevance in the field of forensic facial comparison, where despite unfamiliar facial identification being complex and unreliable (Bruce *et al.*, 2001; Burton *et al.*, 2010), experts have been demonstrated to perform notably better than members of the public (Wilkinson and Evans, 2009). Notwithstanding the importance of experts in facial comparison, according to analyses conducted in 2017 by the Criminal Record Centre (CRC) of the South African Police Services (SAPS), only 30 trained specialists in the entirety of SAPS are able to conduct these analyses and testify in court to defend their conclusions (Steyn *et al.*, 2018).

1.6 Approaches to forensic facial comparison

Forensic facial comparison (FFC) can be conducted in one of three primary approaches: facial superimposition, photo-anthropometry, and morphological analysis (Facial Identification Scientific Working Group, 2019; Stephan *et al.*, 2019). These methods all employ a basic level of holistic comparison or holistic review, in the process of comparing faces, but do not exclusively rely on it, as it is highly prone to error when considering unfamiliar faces (Biederman and Kalocsais, 1997; Megreya and Burton, 2006; Rossion, 2008).

1.6.1 Photo-anthropometry

Photo-anthropometric analysis makes use of biometric measurements of various distances, angles and ratios across anthropological landmarks and facial features to compare two or more facial images (Mallett and Evison, 2013; Facial Identification Scientific Working Group, 2019; Stephan *et al.*, 2019). A variety of photogrammetry approaches and landmarks have been used in photo-anthropometry both in two-dimensional (2D) (e.g., Flores *et al.*, 2019) and three-dimensional (3D) images (e.g., Wong *et al.*, 2008), with the majority of them relying on Farkas' atlas (1994) of craniofacial anthropometric landmarks and features. Following the development of these landmarks, the field of anthropometry has progressed significantly. This is particularly evident with the development of digital anthropometry, where anthropometric landmarks can be automatically placed as coordinates on digital image data and measurements automatically calculated based on these coordinate networks (Kukharev and Kaziyeva, 2020). While digital anthropometry has numerous applications in industrial and market areas, these automated landmarking systems suffer from the same shortcomings as automated facial recognition of relying on standardised images captured under controlled conditions and is prone to error in realistic scenarios (e.g., Urbanová, 2016). 2D photo-anthropometry is considered the most cost-effective approach for quantification of the facial morphology and is of great use in anthropology, however, it is underperforms in accuracy in comparison to 3D photo-anthropometry (Anas *et al.*, 2019). Despite the promise demonstrated by 3D photo-anthropometry, application in the forensic context is complex due to the lack of 3D surveillance data in realistic scenarios.

Unfortunately, photo-anthropometry has been identified as not able to meet admissibility criteria required in a court of law (Lucas *et al.*, 2016), due to its high error rates and potential to mislead, as demonstrated in multiple studies (Kleinberg *et al.*, 2007; Evison *et al.*, 2010; Moreton and Morley, 2011). This poor performance of photo-anthropometry is enhanced by its significant increase in susceptibility to error in sub-optimal images (Kleinberg *et al.*, 2007; Evison *et al.*, 2010; Moreton and Morley, 2011; Lucas *et al.*, 2016). Conditions rendering images suboptimal for use in photo-anthropometry include low resolution, non-corresponding facial views or camera angles, inconsistent lighting conditions, or motion blur (Kleinberg *et al.*, 2007; Moreton and Morley, 2011; Facial Identification Scientific Working Group, 2019). These are all factors that affect the visibility of facial features and ability to identify landmarks accurately (Stavrianos *et al.*, 2012). In addition, to be able to acquire photogrammetric information from the image data, knowledge of the focal length, lens distortion and subject

distance is required (Facial Identification Scientific Working Group, 2019). Under all these requirements and considerations, in realistic conditions such as when using CCTV recordings, it may be impossible to extract suitable images to confidently ascertain similarity or dissimilarity in a forensic facial comparison (Facial Identification Scientific Working Group, 2019).

1.6.2 Photographic superimposition

An alternative method of facial comparison is facial photographic superimposition. Facial photographic superimposition entails overlaying two scaled and rotated facial images to visually compare correspondence (Stavrianos *et al.*, 2012; Mallett and Evison, 2013; Facial Identification Scientific Working Group, 2019). This method has been employed since the 1990s with moderate success yet with multiple limitations and requirements (Vanezis and Brierley, 1996). Some of these concerns include the requirement for a correspondence in face orientation within a tolerable threshold, the effect of fading or wiping an image over the other, as well as high quality images and equipment to conduct the comparisons (Vanezis and Brierley, 1996). A major concern is the strong bias of innate human face recognition to blend superimposed images together during wiping or fading, providing the illusion of correspondence, overall reducing the accuracy of facial comparisons (Vanezis and Brierley, 1996; Strathie and McNeill, 2016). In light of these concerns, superimposition has been widely recommended strictly for exclusion purposes and not for positive identification (Davis *et al.*, 2010; Stavrianos *et al.*, 2012; Facial Identification Scientific Working Group, 2019). Later advances extrapolated the superimposition approach to compare 2D facial images to 3D facial images in an attempt to resolve incongruencies due to camera angle and facial view (e.g. Yoshino, 2005). Although this approach was also able to correct for perspective distortion provided the image capture distance was known (Yoshino, 2005), obtaining 3D facial scans of prospective matches would be a complex task in most forensic settings (Stephan *et al.*, 2019).

1.6.3 Morphological analysis

The final approach to facial comparison is morphological analysis. Morphological analysis involves the direct comparison of individual facial features and components between two or more faces, with a final conclusion relating to similarity or dissimilarity of the faces reliant on subjective assessment, evaluation and interpretation of individual observations (Facial Identification Scientific Working Group, 2019). The facial features in question can vary based on the specific approach or feature list employed, but feature lists usually include overall face

composition and structure, individual anatomical feature components (e.g. forehead shape, ear helix, mouth vermillion borders, *etc*.) as well as distinguishing characteristics such as face marks, scars, piercings, and tattoos (e.g. Facial Identification Scientific Working Group, 2018, 2019). The morphological analysis method is designed and intended to be applied during the ACE-V process of facial comparison. An important distinction between the FISWG approach to morphological analysis and older attempts at feature-based comparisons are classification systems or schemes. These classification schemes require an observer to categorically assign a generic and descriptive quality to facial features (e.g. square face, straight nose *etc*.). This is a distinct process from the FISWG recommended approach of subjectively describing the facial appearance of the compared faces based on similarities and dissimilarities (Facial Identification Scientific Working Group, 2019). Classification schemes are not considered best practice (Facial Identification Scientific Working Group, 2019) due to them being prone to high inter-observer errors (Vanezis *et al.*, 1996; Ritz-Timme *et al.*, 2011a). A similar concern was raised with fingerprint analysis, where poor inter-rater consistency was believed to result from lacking objective and quantifiable information, particularly in data of suboptimal quality (Dror *et al.*, 2011). Classification schemes such as these require testing for population homogeneity, which only has been conducted by some studies (Roelofse *et al.*, 2008; Ritz-Timme *et al.*, 2011a). Some features may present with high levels of prevalence, in a given population, which may lead to faces being classified similarly under the exact same categories resulting in matches assigned based on classification schemes to be erroneous and inconsistent (Ritz-Timme *et al.*, 2011a). On the opposite end of the spectrum, another valid concern with this approach to morphological analysis is that instances may arise where classification of a facial feature into a specific category may be hard to impossible (Valentine and Davis, 2015). It is important to note however, that these examples of classification systems relied on simple classifications with minimal description or definition (e.g. (Vanezis *et al.*, 1996; Ritz-Timme *et al.*, 2011a), yet were still found to perform better in identification than with some form of feature guidance (Ritz-Timme *et al.*, 2011a).

In contrast to the concerns raised by some researchers, facial feature descriptors and ratings have recently been associated with notably higher accuracies in both memory and non-memory facial matching tasks (Towler *et al.*, 2017; Megreya and Bindemann, 2018). Megreya and Bindemann (2018) identified that even simple directions to pay attention to individual facial features improved face matching accuracy of inexperienced analysts. In accordance with this, Towler *et al.* (2017) found that combined feature similarity ratings of features not only

improved accuracy but combined multiple analysts scores would increase the power of identity discrimination to 99%. This increase in discriminatory ability was even more pronounced with experts, where a single expert provided an increase in performance equivalent to 7 inexperienced observers' combined ratings (Towler *et al.*, 2017). As a result, the importance of expertise is further reinforced as previously demonstrated by Wilkinson and Evans (2009) who identified a notably better performance when experts were performing facial identification tasks compared to the general public. However, experts were still found likely to falsely accept individuals with similar features as the target with moderate support (Wilkinson and Evans, 2009). This would suggest that even among experts there is a real risk of mistaken identification that could lead to wrongful conviction (Valentine, 2006). Expertise, while still prone to false classifications in approximately 7% of cases, has also recently been shown to play a role not only in considered judgements, such as morphological analysis, but even in intuitive facial comparison limited to short exposure to the image data (White *et al.*, 2015).

1.7 Morphological analysis as best practice

Among the various approaches to FFC discussed above, the Facial Identification Scientific Working Group (FISWG) (https://fiswg.org/index.htm) recommends facial comparison *via* morphological analysis as the best practice method for forensic facial identification (Facial Identification Scientific Working Group, 2012, 2019). The FISWG was established as the primary working group involved in standard and guideline development for the field of craniofacial identification. This group is part of the Scientific Committees for Forensic Sciences operating under the United States National Institute for Sciences and Technology (*Facial Identification Scientific Working Group: About Scientific Working Groups*, 2019). The Scientific Committees for Forensic Sciences operate in collaboration and support with over 500 forensic science practitioners and other experts for the reinforcement of forensic science in the United States (*Facial Identification Scientific Working Group: About Scientific Working Groups*, 2019). The Working Groups are well-known for their contributions in a multitude of forensic subject areas including, for example, DNA analysis (https://www.swgdam.org), firearm and toolmark analysis (https://afte.org), as well as digital and multimedia evidence acquisition and analysis (https://www.swgde.org). The FISWG recommends that photo-anthropometry should not be used for facial image comparison, while suggesting that superimposition should only be employed in combination with morphological analysis (Facial Identification Scientific Working Group, 2019).

To this end, the FISWG has developed and published its own extensive facial feature list for use in morphological analysis, including a total of 18 broad facial components and an additional component labelled as "Other" for uncategorised components or features (Facial Identification Scientific Working Group, 2013, 2018). These facial components include the following: skin, face and head outline, face and head composition, hair, forehead, eyebrows, eyes, cheeks, nose, ears, mouth, chin/jawline, neck, facial hair, facial lines, scars, facial marks, and alterations (Facial Identification Scientific Working Group, 2013, 2018). Each facial component is further individually discussed, with many specific characteristics to analyse (Facial Identification Scientific Working Group, 2013, 2018). These characteristics are also associated to common descriptive terminology (e.g. portrait contour of the cranial vault, arch of the eyebrow) in order to highlight the specific details to be analysed, almost as a rudimentary checklist, and to aid facial comparison analysts in the descriptive process of reporting. These descriptors are extensive in nature and rely on the anatomical detail of the face, often making use of anthropological terminology. Other feature lists exist that were previously developed as classification systems and atlases (Vanezis *et al.*, 1996; Ohlrogge *et al.*, 2008, 2009; Roelofse *et al.*, 2008; Ritz-Timme *et al.*, 2011a; Ritz-Timme *et al.*, 2011b) or have been broadly considered (Porter and Doran, 2000). However, these lists are not as extensive as that of FISWG, with the most extensive and freely available one including a total of 43 facial characteristics (Ritz-Timme *et al.*, 2011a; Ritz-Timme *et al.*, 2011b). The FISWG feature list is the most thorough list available, and includes over 130 facial component characteristics and over 290 characteristic descriptors (Facial Identification Scientific Working Group, 2018). With the FISWG feature list, in depth morphological comparison of faces can thus be conducted.

Although facial comparison and morphological analysis specifically are widely accepted by its practitioners worldwide (Wilkinson and Evans, 2009; Houlton and Steyn, 2018; Steyn *et al.*, 2018; Facial Identification Scientific Working Group, 2019), there is no extensive scientific evidence testing the accuracy of the method, and as such, it does not conform to the requirements of admissibility (Mallett and Evison, 2013). This is of particular importance when considering different real-world scenarios, and not simply an optimal standardised context. An optimal setting includes availability of mugshots and other standardised photographic imagery obtained for the purpose of facial image comparison. While these are useful for the baseline image comparison, they do not adequately simulate different realistic scenarios with recordings from multiple angles, obstructions and varied facial expressions. The lack of scientific trials

for facial comparison by morphological analysis, particularly with the use of the FISWG feature list, contributes to the risk of the expert report, submitted as evidence to court, to be discarded as inadmissible. Although previously the deliberations of facial identification experts have been deemed "admissible and frequently of value" (Court of Appeal, 2003 cited in: Wilkinson and Evans, 2009, p. 192), concerns have been raised in the past in the forensic community about many established forensic methods (Cunliffe, 2013; Curry, 2015). Forensic facial identification has not escaped this scrutiny of forensic methods relating to admissibility to court of forensic evidence (Jackson, 2016).

The fact that forensic facial identification is commonly employed despite being untested becomes progressively problematic due to the continuing increase in CCTV recordings available to police services (Jackson, 2016), which leads to intensive use of facial comparison with no thorough investigation into its reliability. In Australia, facial comparison experts presenting evidence in the legal system are only expected to describe the similarities and dissimilarities of the faces in question, which may result in strongly suggestive language based on prejudice and opinion from the expert's perspective (Valentine and Davis, 2015). However, in England and Wales facial comparison experts present their evidence through scales of support, specifically using the Bromby scale of support (Bromby, 2006), to reinforce the confidence level of their conclusions (Valentine and Davis, 2015). Such scales force an expert to conclude whether the two compared images belong to the same person or not regardless of image quality (Valentine and Davis, 2015). In addition, this scale is inherently arbitrary with minimal definition of each step of the scale and would require testing before application. The risk of false positive and false negative reports are of prime concern, as a false negative identification would fail to recognise a suspect as the perpetrator and might enhance the defence attorney's motivation for their innocence, while a false positive would risk wrongful conviction (Valentine and Davis, 2015). To this end, South African Police Services' experts use a five point scale that reflects the ability as well as the confidence of a specific conclusion, allowing for inconclusive analyses to be scored as such, taking into account image quality and other factors (Steyn *et al.*, 2018).

In the South African legal system, expert reports can only be considered as evidence if all the necessary facial identification criteria utilised to obtain the expert's conclusion are included and informed (Faurie, 2000). As there is little to no testing of the morphological analysis, once all the criteria are explained in court, the perceived subjectivity of the method may be rebutted to render the evidence inadmissible. Thus, a more rigorous scientific approach to facial

comparison by morphological analysis is needed. This is of particular importance in consideration of the varied conditions of image-based evidence available to conduct facial comparison.

1.8 Limitations of CCTV in facial identification

Advancements in the computing industry and the reduced cost of electronics, has driven a rise in personal, public, and commercial use of CCTV systems (Norris *et al.*, 2002). This increased deployment of CCTV systems allows for superior monitoring of private and public property, and surveillance of potential criminal activity, which directly leads to an increase of available image-based data for criminal investigations. However, the large majority of image data used in criminal investigations is of low to medium quality (Gibelli *et al.*, 2016). These suboptimal images are often due to poor recording conditions, outdated surveillance systems, and optical distortions (Gibelli *et al.*, 2016). Furthermore, due to the variation in recording angles, altered proportions, and varied distances resulting from most unstandardised CCTV installations, facial images tend to be hard to compare to mugshots or other standardised photographs. Although general awareness around these very real limitations of surveillance data exists, few studies have considered their implication in facial comparison accuracy (Burton *et al.*, 1999; Lee *et al.*, 2009; Bindemann *et al.*, 2013; Ritchie *et al.*, 2018). With inconsistent recording conditions and poor quality, correct facial identification ability is known to decrease (Fysh and Bindemann, 2018; Kramer *et al.*, 2019). This is particularly evident in the marked variation in individual accuracy across multiple analysts (Burton *et al.*, 2010). The negative impact of image quality in face matching accuracy can be further compounded by the common overestimation of an analyst's personal ability to identify faces (Bindemann *et al.*, 2014).

1.9 Image quality

Poor quality footage was found to affect face matching ability notably in both members of the public and experts, with accuracies of 66% to 67% respectively, and overall high false positive rates (Lee *et al.*, 2009). A common form of a poor-quality artefact affecting facial matching is image pixelation (Bindemann *et al.*, 2013; Ritchie *et al.*, 2018). When observing face matching ability in pixelated images, Ritchie *et al.* (2018) demonstrated accuracies ranging between 40% and 65%, while Bindemann *et al.* (2013) found a decrease in accuracy from approximately 90% in a high-quality image sample to as low as 48% in their poorest quality images.

As such, an imminent challenge in morphological comparison remains image quality. Poor image quality decreases the amount of information that can be extracted from images. This is particularly noteworthy across different types of CCTV systems. Traditional analogue CCTV systems record in low resolution, often monochrome imagery, which completely removes one of the first components of morphological analysis (colour), while producing low quality images with low clarity, compounding the difficulty of identification (Burton *et al.*, 1999). Current Internet Protocol (IP) CCTV systems have full colour images and a substantially better resolution that provide an analyst with better resources to compare to a database (Surette, 2005).

Additionally, lighting conditions can pose further challenges in recording optimal footage. Overexposure of a subject can gravely hinder the definition of details in a face (Damjanovski, 2014). This may occur due to the position of the sun at peak times, in outdoor locations, or in instances where multiple light sources or reflective materials are placed near the camera. These extreme lighting conditions can overexpose the target and create artificial boundaries on the viewed objects, altering their appearance (Damjanovski, 2014). On the opposite side of the spectrum, low light conditions will also obscure facial appearance in the recordings. While this can be counteracted with lowlight/infrared cameras, these cameras rely on an infrared source of light that could itself overexpose the subject in close proximity and is particularly susceptible to reflective materials (Damjanovski, 2014).

Detecting an optimally visible face in CCTV footage may not always be possible. In order to inspect facial features for facial image comparison, according to conservative European standards (Damjanovski, 2014), a minimum of the top quarter of an individual, basically from the bust and above, must be included on the screen height (British Security Industry, 2014). In terms of resolution, the face would then have to represent a minimum of 1000 pixels per meter of screen height (British Security Industry, 2014). These dimensions effectively translate to 25.4 pixels per inch, a unit comparable to the commonly used dots per inch (dpi). As such, for each inch of the face represented in an image, there should be a minimum of 25.4 pixels constituting it. To achieve this minimum pixel density at a range of 5 m, for example, a focal length of 4.2 mm is necessary on a 1/2" high-definition CCTV camera, according to Damjanovski's (2014) formula. A 1/3" standard definition camera would require a focal length of 8.5 mm at a 5 m distance (Damjanovski, 2014). Focal lengths such as these, or even greater ones, narrow the field of view of a particular lens (Damjanovski, 2014). As such, cameras installed specifically to monitor wide areas will not be effective at capturing a face for facial

examination purposes (Damjanovski, 2014). Unfortunately, this is the most common scenario, as the majority of individuals installing CCTV systems attempt to cut implementation costs without realising the implications with regard to law enforcement. Most surveillance cameras are also mounted well above eye level height, at heights around to 2.5 m and 3 m and may result in sub-optimal angles of incidence and focal lengths of capture (Damjanovski, 2014), hindering facial comparison.

1.10 Disguises and facial identification

In addition to sub-optimal images, disguises such as scarves, sunglasses, and hats can prevent a camera from capturing a complete face. Despite the obvious and well know effect of disguises on facial identification (Patterson and Baddeley, 1977; Ryder *et al.*, 2015) only a handful of studies consider them in human based face recognition (Terry, 1993; Lee *et al.*, 2009; Mansour *et al.*, 2012; Megreya *et al.*, 2012; Righi *et al.*, 2012). Glasses and hats are considered some of the less conspicuous and most common everyday disguises used to conceal facial appearance (Terry, 1993; Henderson *et al.*, 2001; Davis *et al.*, 2012). While varying effects for the various types of disguises have been described, no studies attempted a truly morphological approach to their experiments. Irrespective of this, greater error rates were identified for brimmed caps compared to other samples (Henderson *et al.*, 2001), while brimless caps and glasses appeared to have lesser effects on matching ability, depending on the methodological approach taken by each study (Lee *et al.*, 2009; Righi *et al.*, 2012). With the exception of one study (Davis and Valentine, 2009), both glasses and hats, irrespective of type, decreased the capacity of observers to correctly match facial images (Terry, 1993; Henderson *et al.*, 2001; Lee *et al.*, 2009; Mansour *et al.*, 2012; Righi *et al.*, 2012; Kemp *et al.*, 2016; Davis and Tamonyte, 2017). As most of these studies involved memory tasks and on occasion live participants as opposed to digital images, these results are hard to compare and particularly difficult to extrapolate to a forensic setting, requiring thorough morphological analysis with an extensive feature list. Hence, research examining the effects of disguises is also crucial to optimising and determining the efficacy of facial identification from CCTV recordings.

1.11 The need for facial image databases

To test morphological analysis across these various iterations of realistic conditions presented above, a large sample of corresponding photographs and CCTV recordings are required. A total of about 27 reputable face databases are available from various institutions or companies (e.g.

Martinez and Benavente, 1998; Phillips *et al.*, 2000, 2006; Sim *et al.*, 2003; Calvo and Lundqvist, 2008; Gao *et al.*, 2008; Milborrow *et al.*, 2008; Tottenham *et al.*, 2009; Grgic *et al.*, 2011; Kostinger *et al.*, 2011; Sagonas *et al.*, 2015; Karam and Zhu, 2015; Ma *et al.*, 2015). However, most of these databases are usually specific to their original purpose and designed under niche criteria not necessarily suitable for different research approaches (Gross, 2005). As a result, the various databases are not always comparable, often highlighting varieties of different image capture conditions and criteria such as variations in facial expressions (e.g. Martinez and Benavente, 1998; Phillips *et al.*, 2000; Tottenham *et al.*, 2009), facial poses (e.g. Grgic *et al.*, 2011; Sagonas *et al.*, 2015), hairstyles (e.g. Huang *et al.*, 2007; Kostinger *et al.*, 2011), disguises (e.g. Gourier *et al.*, 2004; Gao *et al.*, 2008), and lightning (e.g. Marszalec *et al.*, 2000; Ma *et al.*, 2015). Many of these databases consist of a limited amount of unique faces (e.g. Belhumeur and Kriegman, 1997; Georghiades *et al.*, 2001) and others consist of low resolution images (e.g. Phillips *et al.*, 2000; Huang *et al.*, 2007; Karam and Zhu, 2015). Out of these datasets, only one includes surveillance recordings in conjunction with corresponding high resolution photographic imagery (Grgic *et al.*, 2011) and only a single other African database is available which makes use of low-resolution imagery with poor lighting (Milborrow *et al.*, 2008).

1.12 Aims and objectives

This study aimed to determine the accuracy and reliability of forensic facial comparison via morphological analysis in photographic and varied CCTV conditions. To this end, a twofold approach was required with the following two major aims and subsequent objectives:

1. Firstly, a standardised yet actualistic large-scale database of corresponding facial photographs and CCTV recordings was established.
    a. Standardised and unstandardised photographs of participants were captured.
    b. Corresponding CCTV recordings of participants were collected on university premises via the pre-existing CCTV systems, under various conditions, including different quality cameras and with disguises (glasses, caps, etc.).
2. Secondly, a practical series of experiments needed to be conducted in order to confirm the reliability and accuracy of facial comparison done under optimal and suboptimal image capture circumstances, in order to explore the validity and error rates of

morphological analysis using the most extensive feature list developed by the FISWG. For this purpose, the experiments as outlined below were conducted.

a. Validation of morphological analysis was tested in a photographic and standard CCTV set-up.

b. Variation in morphological analysis accuracy and error rates across high-quality digital and low-quality analogue CCTV data were tested.

c. The effects of facial disguises, (brimmed caps and sunglasses) on morphological analysis accuracy and error rates were tested.

References – Integrated narrative

Adamovic, M. (2021) Crime Index by Country 2021, Numbeo.com. Available at: https://www.numbeo.com/crime/rankings_by_country.jsp (Accessed: 31 January 2021).

Anas, I. Y., Bamgbose, B. O. and Nuhu, S. (2019) 'A comparison between 2D and 3D methods of quantifying facial morphology', Heliyon, 5(6), p. e01880. doi: 10.1016/j.heliyon.2019.e01880.

Ashby, M. P. J. (2017) 'The Value of CCTV Surveillance Cameras as an Investigative Tool: An Empirical Analysis', European Journal on Criminal Policy and Research, 23(3), pp. 441–459. doi: 10.1007/s10610-017-9341-6.

Belhumeur, P. N. and Kriegman, D. J. (1997) 'Eigenfaces vs . Fisherfaces : Recognition Using Class Specific Linear Projection', in IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, pp. 711–720.

Bell, A. (2018) 'Crime scene photography in England, 1895-1960', Journal of British Studies, 57(1), pp. 53–78. doi: 10.1017/jbr.2017.182.

Bertillon, A. and McClaughry, R. W. (1896) Signaletic Instructions Including the Theory and Practice of Anthropometrical Identification. Edited by R. W. McClaughry. Harvard: Warner Company. Available at: https://books.google.co.za/books?hl=en&lr=&id=KG1Nl2ytKcIC&oi=fnd&pg=IA11&ots=vvRhWQEANc&sig=0cR3JLidmuEMyLXZiU5dkunBP3g&redir_esc=y#v=onepage&q&f=false (Accessed: 12 February 2021).

Biederman, I. and Kalocsais, P. (1997) 'Neurocomputational bases of object and face recognition', Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences. Edited by J. Anderson, H. B. Barlow, and R. L. Gregory, 352(1358), pp. 1203–1219. doi: 10.1098/rstb.1997.0103.

Bindemann, M. et al. (2013) 'The effect of image pixelation on unfamiliar-face matching', Applied Cognitive Psychology, 27(6), pp. 707–717. doi: 10.1002/acp.2970.

Bindemann, M., Attard, J. and Johnston, R. A. (2014) 'Perceived ability and actual recognition accuracy for unfamiliar and famous faces', Cogent Psychology, 1(1). doi: 10.1080/23311908.2014.986903.

Boyce, M. A., Lindsay, D. S. and Brimacombe, C. A. E. (2008) 'Investigating investigators: Examining the impact of eyewitness identification evidence on student-investigators.', Law and Human Behavior, 32(5), pp. 439–453. doi: 10.1007/s10979-007-9125-5.

British Security Industry (2014) Video surveillance systems for use in security applications: system requirements - general.

Bromby, M. (2006) 'CCTV and expert evidence: addressing the reliability of new sciences', Archbold News, (9), pp. 6–9. Available at: https://ssrn.com/abstract=1551791.

Bruce, V. et al. (2001) 'Matching identities of familiar and unfamiliar faces caught on CCTV images.', Journal of Experimental Psychology: Applied, 7(3), pp. 207–218. doi: 10.1037/1076-898X.7.3.207.

Burton, A. M. et al. (1999) 'Face recognition in poor-quality video: evidence from security surveillance', Psychological Science, 10(3), pp. 243–248. doi: 10.1111/1467-9280.00144.

Burton, A. M., White, D. and McNeill, A. (2010) 'The Glasgow Face Matching Test', Behavior Research Methods, 42(1), pp. 286–291. doi: 10.3758/BRM.42.1.286.

Calvo, M. G. and Lundqvist, D. (2008) 'Facial expressions of emotion (KDEF): Identification under different display-duration conditions', Behavior Research Methods, 40(1), pp. 109–115. doi: 10.3758/BRM.40.1.109.

Cunliffe, E. (2013) 'Independence, reliability and expert testimony in criminal trials', Australian Journal of Forensic Sciences, 45(3), pp. 284–295. doi: 10.1080/00450618.2013.784358.

Curry, B. M. (2015) 'Science in court: challenging the value of expert evidence', (September), pp. 84–85.

Damjanovski, V. (2014) CCTV from light to pixels. 3rd edn, CCTV from light to pixels. 3rd edn. Edited by V. Damjanovski. Oxford: Elsevier.

Davis, J. P. and Tamonyte, D. (2017) 'Masters of disguise: Super-recognisers' superior memory for concealed unfamiliar faces', in 2017 Seventh International Conference on Emerging Security Technologies (EST). IEEE, pp. 44–49. doi: 10.1109/EST.2017.8090397.

Davis, J. P. and Valentine, T. (2009) 'CCTV on trial: Matching video images with the defendant in the dock', Applied Cognitive Psychology, 23(4), pp. 482–505. doi: 10.1002/acp.1490.

Davis, J. P., Valentine, T. and Davis, R. E. (2010) 'Computer assisted photo-anthropometric analyses of full-face and profile facial images', Forensic Science International, 200(1–3), pp. 165–176. doi: 10.1016/j.forsciint.2010.04.012.

Davis, J. P., Valentine, T. and Wilkinson, C. (2012) 'Facial image comparison', in Wilkinson, C. and Rynn, C. (eds) Craniofacial Identification. 1st edn. New York: Cambridge University Press, pp. 136–153. doi: 10.1017/CBO9781139049566.015.

Dodd, V. (2018) 'UK police use of facial recognition technology a failure, says report', The Guardian, 15 May. Available at: https://www.theguardian.com/uk-news/2018/may/15/uk-police-use-of-facial-recognition-technology-failure.

Dror, I. E. et al. (2011) 'Cognitive issues in fingerprint analysis: Inter- and intra-expert consistency and the effect of a "target" comparison', Forensic Science International, 208(1–3), pp. 10–17. doi: 10.1016/j.forsciint.2010.10.013.

Duncan, J. (2018) 'How CCTV surveillance poses a threat to privacy in South Africa', The Conversation, pp. 1–3. Available at: https://www.iol.co.za/saturday-star/news/how-cctv-surveillance-poses-a-threat-to-privacy-in-south-africa-15342438.

Evison, M. et al. (2010) 'Key parameters of face shape variation in 3D in a large sample', Journal of Forensic Sciences, 55(1), pp. 159–162. doi: 10.1111/j.1556-4029.2009.01213.x.

Facial Identification Scientific Working Group: About Scientific Working Groups (2019). Available at: https://fiswg.org/about_swgs.html (Accessed: 12 February 2021).

Facial Identification Scientific Working Group (2012) FISWG Guidelines for Facial Comparison Methods. Version 1.0 - 2012-02-02. Available at: https://www.fiswg.org/FISWG_GuidelinesforFacialComparisonMethods_v1.0_2012_02_02.pdf. (Accessed: 16 March 2021).

Facial Identification Scientific Working Group (2013) 'Facial Image Comparison Feature List for Morphological Analysis'. Version 1.0 2013.11.22. Available at: https://fiswg.org/FISWG_1to1_Checklist_v1.0_2013_11_22_ARCHIVED.pdf (Accessed: 16 March 2021).

Facial Identification Scientific Working Group (2018) Facial Image Comparison Feature List for Morphological Analysis. Version 2.0 2018.09.11. Available at: https://fiswg.org/FISWG_Morph_Analysis_Feature_List_v2.0_20180911.pdf. (Accessed: 16 March 2021).

Facial Identification Scientific Working Group (2019) Facial Comparison Overview and Methodology Guidelines. Version 1.0 2019.10.25. Available at: https://fiswg.org/fiswg_facial_comparison_overview_and_methodology_guidelines_V1.0_20191025.pdf. (Accessed: 16 March 2021).

Faigman, D. L. (2008) 'Anecdotal Forensics, Phrenology, and Other Abject Lessons from the History of Science', Hastings Law Journal, 59(5), pp. 979–1000.

Farkas, L. G. (1994) Anthropometry of the head and face. Second. Edited by L. G. Farkas. New York, USA: Raven Press.

Faurie, A. (2000) The admissibility and evaluation of scientific evidence in court. University of South Africa.

Flores, M. R. P. et al. (2019) 'Comparative Assessment of a Novel Photo-Anthropometric Landmark-Positioning Approach for the Analysis of Facial Structures on Two-Dimensional Images', Journal of Forensic Sciences, 64(3), pp. 828–838. doi: 10.1111/1556-4029.13935.

Fysh, M. C. and Bindemann, M. (2018) 'The Kent Face Matching Test', British Journal of Psychology, 109(2), pp. 219–231. doi: 10.1111/bjop.12260.

Gao, W. et al. (2008) 'The CAS-PEAL large-scale chinese face database and baseline evaluations', in IEEE Transactions on Systems, Man, and Cybernetics, pp. 149–161. doi: 10.1109/TSMCA.2007.909557.

Georghiades, A. S., Belhumeur, P. N. and Kriegman, D. J. (2001) 'From few to many: Illumination cone models for face recognition under variable lighting and pose', in IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 643–660. doi: 10.1109/34.927464.

Gibelli, D. et al. (2016) 'The identification of living persons on images: A literature review', Legal Medicine, 19, pp. 52–60. doi: 10.1016/j.legalmed.2016.02.001.

Gill, M. and Spriggs, A. (2005) Assessing the impact of CCTV: Home Office Research Study, Home office Research Study 296. Available at: https://techfak.uni-

bielefeld.de/~iluetkeb/2006/surveillance/paper/social_effect/CCTV_report.pdf. (Accessed: 16 March 2021).

Goold, B., Loader, I. and Thumala, A. (2013) 'The banality of security: The curious case of surveillance cameras', British Journal of Criminology, 53(6), pp. 977–996. doi: 10.1093/bjc/azt044.

Gourier, N., Hall, D. and Crowley, J. L. (2004) 'Estimating face orientation from robust detection of salient facial structures', in International Workshop on Visual Observation of Deicitic Gestures, pp. 17–25. doi: 10.1.1.329.1780.

Grgic, M., Delac, K. and Grgic, S. (2011) 'SCface - Surveillance cameras face database', Multimedia Tools and Applications, 51(3), pp. 863–879. doi: 10.1007/s11042-009-0417-2.

Gross, R. (2005) 'Face Databases', in Li, S. and Jain, A. (eds) Handbook of Face Recognition. New York: Springer-Verlag, pp. 301–327. doi: 10.1007/0-387-27257-7_14.

Grother, P., Ngan, M. and Hanaoka, K. (2018) Ongoing face recognition vendor test (FRVT) part 2: identification. Gaithersburg, MD. doi: 10.6028/NIST.IR.8238.

Grother, P., Ngan, M. and Hanaoka, K. (2019) Face recognition vendor test part 3: demographic effects, Nistir 8280. Gaithersburg, MD. doi: 10.6028/NIST.IR.8280.

Henderson, Z., Bruce, V. and Burton, A. M. (2001) 'Matching the faces of robbers captured on video', Applied Cognitive Psychology, 15(4), pp. 445–464. doi: 10.1002/acp.718.

Houlton, T. M. R. and Steyn, M. (2018) 'Finding Makhubu: A morphological forensic facial comparison', Forensic Science International, 285, pp. 13–20. doi: 10.1016/j.forsciint.2018.01.022.

Huang, G. B. et al. (2007) Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments, University of Massachusetts, Amherst Technical Report. doi: https://hal.inria.fr/inria-00321923.

Jackson, A. (2016) 'The Admissibility of Identification Evidence Made on the Basis of Recognition from Photographs Taken at a Crime Scene', The Journal of Criminal Law, 80(4), pp. 234–236. doi: 10.1177/0022018316659958.

Jäger, J. (2001) 'Photography: a means of surveillance? Judicial photography, 1850 to 1900', Crime, Histoire & Sociétés, 5(1), pp. 27–51. doi: 10.4000/chs.1056.

Jain, A. K., Klare, B. and Park, U. (2012) 'Face Matching and Retrieval in Forensics Applications', IEEE Multimedia, 19(1), pp. 20–20. doi: 10.1109/MMUL.2012.4.

Karam, L. J. and Zhu, T. (2015) 'Quality labeled faces in the wild (QLFW): a database for studying face recognition in real-world environments', Human Vision and Electronic Imaging XX, 9394, p. 93940B. doi: 10.1117/12.2080393.

Kaur, P. et al. (2020) 'Facial-recognition algorithms: A literature review', Medicine, Science and the Law, 60(2), pp. 131–139. doi: 10.1177/0025802419893168.

Kemp, R. I. et al. (2016) 'Improving Unfamiliar Face Matching by Masking the External Facial Features', Applied Cognitive Psychology, 30(4), pp. 622–627. doi: 10.1002/acp.3239.

Kleinberg, K. F. and Siebert, J. P. (2012) 'A study of quantitative comparisons of photographs and video images based on landmark derived feature vectors', Forensic Science International, 219(1–3), pp. 248–258. doi: 10.1016/j.forsciint.2012.01.014.

Kleinberg, K. F., Vanezis, P. and Burton, A. M. (2007) 'Failure of anthropometry as a facial identification technique using high-quality photographs', Journal of Forensic Sciences, 52(4), pp. 779–783. doi: 10.1111/j.1556-4029.2007.00458.x.

Kostinger, M. et al. (2011) 'Annotated Facial Landmarks in the Wild: a large-scale, real-world database for facial landmark localization', in 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops). IEEE, pp. 2144–2151. doi: 10.1109/ICCVW.2011.6130513.

Kramer, R. S. S., Mohamed, S. and Hardy, S. C. (2019) 'Unfamiliar Face Matching with Driving Licence and Passport Photographs', Perception, 48(2), pp. 175–184. doi: 10.1177/0301006619826495.

Kukharev, G. A. and Kaziyeva, N. (2020) 'Digital Facial Anthropometry: Application and Implementation', Pattern Recognition and Image Analysis, 30(3), pp. 496–511. doi: 10.1134/S1054661820030141.

Lee, W.-L. et al. (2009) 'Matching unfamiliar faces from poor quality closed-circuit television (CCTV) footage: An evaluation of the effect of training on facial identification ability', AXIS Online Journal of Centre for Anatomy and Human Identification, 1(1), pp. 19–28.

Lindegaard, M. R. and Bernasco, W. (2018) 'Lessons Learned from Crime Caught on Camera', Journal of Research in Crime and Delinquency, 55(1), pp. 155–186. doi: 10.1177/0022427817727830.

Lucas, T., Kumaratilake, J. and Henneberg, M. (2016) 'Metric Identification of the Same People from Images: How Reliable Is It?', Journal of Anthropology, 2016, pp. 1–10. doi: 10.1155/2016/1760313.

Ma, D. S., Correll, J. and Wittenbrink, B. (2015) 'The Chicago face database: A free stimulus set of faces and norming data', Behavior Research Methods, 47(4), pp. 1122–1135. doi: 10.3758/s13428-014-0532-5.

Mallett, X. and Evison, M. P. (2013) 'Forensic Facial Comparison: Issues of Admissibility in the Development of Novel Analytical Technique', Journal of Forensic Sciences, 58(4), pp. 859–865. doi: 10.1111/1556-4029.12127.

Mansour, J. K. et al. (2012) 'Impact of disguise on identification decisions and confidence with simultaneous and sequential lineups.', Law and Human Behavior, 36(6), pp. 513–526. doi: 10.1037/h0093937.

Marszalec, E. et al. (2000) 'Physics-based face database for color research', Journal of Electronic Imaging, 9(1), pp. 32–28. doi: 10.1117/1.482722.

Martinez, A. M. and Benavente, R. (1998) 'The AR Face Database CVC Tech. Report #24', 24. Available at: http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html. (Accessed: 16 March 2021).

Megreya, A. M. and Bindemann, M. (2018) 'Feature instructions improve face-matching accuracy', PLOS ONE. Edited by P. J. Hills, 13(3), p. e0193455. doi: 10.1371/journal.pone.0193455.

Megreya, A. M. and Burton, A. M. (2006) 'Unfamiliar faces are not faces: Evidence from a matching task', Memory & Cognition, 34(4), pp. 865–876. doi: 10.3758/BF03193433.

Megreya, A. M., Memon, A. and Havard, C. (2012) 'The Headscarf Effect: Direct Evidence from the Eyewitness Identification Paradigm', Applied Cognitive Psychology, 26(2), pp. 308–315. doi: 10.1002/acp.1826.

Milborrow, S., Morkel, J. and Nicolls, F. (2008) 'The MUCT Landmarked Face Database', Pattern Recognition Association of South Africa. doi: http://www.milbo.org/muct/. (Accessed: 16 March 2021).

Mokwena, R. J. (2012) The value of photography in the investigation of crime scenes. University of South Africa. Available at: http://uir.unisa.ac.za/bitstream/handle/10500/7662/dissertation_mokwena_rj.pdf?sequence=1&isAllowed=y. (Accessed: 16 March 2021).

Moreton, R. and Morley, J. (2011) 'Investigation into the use of photoanthropometry in facial image comparison', Forensic Science International, 212(1–3), pp. 231–237. doi: 10.1016/j.forsciint.2011.06.023.

Moyo, S. (2019) Evaluating the use of CCTV surveillance systems for crime cotnrol and prevention: selected case studies from Johannesburg and Tshwane, Gauteng. University of South Africa.

Norris, C., McCahill, M. and Wood, D. (2002) 'The Growth of CCTV: a global perspective on the international diffusion of video surveillance in publicly accessible space', Surveillance & Society, 2(2/3), pp. 110–135. doi: 10.24908/ss.v2i2/3.3369.

Ohlrogge, S. et al. (2008) Anthropological Atlas of Male Facial Features. Edited by S. Ohlrogge et al. Frankfurt, Germany: Verlag für Polizeiwissenschaft.

Ohlrogge, S. et al. (2009) Anthropological Atlas of Female Facial Features. Edited by S. Ohlrogge et al. Frankfurt, Germany: Verlag für Polizeiwissenschaft.

Patterson, K. E. and Baddeley, A. D. (1977) 'When face recognition fails.', Journal of Experimental Psychology: Human Learning and Memory, 3(4), pp. 406–417. doi: 10.1037/0278-7393.3.4.406.

Phillips, P. J. et al. (2000) 'The FERET evaluation methodology for face-recognition algorithms', in IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1090–1104. doi: 10.1109/34.879790.

Phillips, P. J. et al. (2006) 'Preliminary face recognition grand challenge results', in International Conference on Automatic Face and Gesture Recognition, pp. 15–21. doi: 10.1109/FGR.2006.87.

Piza, E. L. et al. (2019) ‘CCTV surveillance for crime prevention: A 40-year systematic review with meta-analysis’, Criminology and Public Policy, 18(1), pp. 135–159. doi: 10.1111/1745-9133.12419.

Porter, G. and Doran, G. (2000) ‘An anatomical and photographic technique for forensic facial identification’, Forensic Science International, 114(2), pp. 97–105. doi: 10.1016/S0379-0738(00)00290-5.

Press Association (2018) ‘Welsh police wrongly identify thousands as potential criminals’, The Guardian, 5 May. Available at: https://www.theguardian.com/uk-news/2018/may/05/welsh-police-wrongly-identify-thousands-as-potential-criminals.

Righi, G., Peissig, J. J. and Tarr, M. J. (2012) ‘Recognizing disguised faces’, Visual Cognition, 20(2), pp. 143–169. doi: 10.1080/13506285.2012.654624.

Ritchie, K. L. et al. (2018) ‘Enhancing CCTV: Averages improve face identification from poor-quality images’, Applied Cognitive Psychology, 32(6), pp. 671–680. doi: 10.1002/acp.3449.

Ritz-Timme S. *et al.* (2011a) ‘A new atlas for the evaluation of facial features: advantages, limits, and applicability’, International Journal of Legal Medicine, 125(2), pp. 301–306. doi: 10.1007/s00414-010-0446-4.

Ritz-Timme S. *et al.* (2011b) ‘Metric and morphological assessment of facial features: A study on three European populations’, Forensic Science International, 207(1–3), pp. 239.e1-239.e8. doi: 10.1016/j.forsciint.2011.01.035.

Roelofse, M. M., Steyn, M. and Becker, P. J. (2008) ‘Photo identification: Facial metrical and morphological features in South African males’, Forensic Science International, 177(2–3), pp. 168–175. doi: 10.1016/j.forsciint.2007.12.003.

Rossion, B. (2008) ‘Picture-plane inversion leads to qualitative changes of face perception’, Acta Psychologica, 128(2), pp. 274–289. doi: 10.1016/j.actpsy.2008.02.003.

Ryder, H., Smith, H. M. J. and Flowe, H. D. (2015) ‘Estimator variables and memory for faces’, in Valentine, T. and Davis, J. P. (eds) Forensic facial identification: theory and practice of identification from eyewitnesses, composites and CCTV. First. John Wiley & Sons, Ltd., pp. 159–183.

Sagonas, C. et al. (2015) '300 Faces In-The-Wild Challenge: database and results', Image and Vision Computing, 47, pp. 3–18. doi: 10.1016/j.imavis.2016.01.002.

Schüler, G. and Obertová, Z. (2020) 'Visual identification of persons: facial image comparison and morphological comparative analysis', in Obertová, Z., Stewart, A., and Cattaneo, C. (eds) Statistics and Probability in Forensic Anthropology. First. London, UK: Elsevier Academic Press, pp. 313–330. doi: 10.1016/B978-0-12-815764-0.00025-3.

Sim, T., Baker, S. and Bsat, M. (2003) 'The CMU Pose, Illumination, and Expression (PIE) database', in IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1615–1618. doi: 10.1109/AFGR.2002.1004130.

South African Police Services (2020) Crime Statistics, Crime Situation in Republic of South Africa Twelve Months (April to Match 2019/2020). doi: 10.4135/9781483340470.n67.

Spaun, N. A. (2009) 'Facial comparisons by subject matter experts: Their role in biometrics and their training', Lecture Notes in Computer Science, 5558, pp. 161–168. doi: 10.1007/978-3-642-01793-3_17.

Speckeis, C. (2011) 'Can ACE-V be validated?', Journal of Forensic Identification, 61(3), pp. 201–209.

Stavrianos, C. et al. (2012) 'Facial mapping: review of current methods', Research Journal of Medical Sciences, 6(2), pp. 77–82. doi: 10.3923/rjmsci.2012.77.82.

Stephan, C. N. et al. (2019) 'An overview of the latest developments in facial imaging', Forensic Sciences Research, 4(1), pp. 10–28. doi: 10.1080/20961790.2018.1519892.

Stevenage, S. V. and Pitfield, C. (2016) 'Fact or fiction: Examination of the transparency, reliability and sufficiency of the ACE-V method of fingerprint analysis', Forensic Science International, 267, pp. 145–156. doi: 10.1016/j.forsciint.2016.08.026.

Steyn, M. et al. (2018) 'Forensic facial comparison in South Africa: State of the science', Forensic Science International, 287, pp. 190–194. doi: 10.1016/j.forsciint.2018.04.006.

Strathie, A. and McNeill, A. (2016) 'Facial Wipes don't Wash: Facial Image Comparison by Video Superimposition Reduces the Accuracy of Face Matching Decisions', Applied Cognitive Psychology, 30(4), pp. 504–513. doi: 10.1002/acp.3218.

Surette, R. (2005) 'The thinking eye', Policing: An International Journal of Police Strategies & Management, 28(1), pp. 152–173. doi: 10.1108/13639510510581039.

Terry, R. L. (1993) 'How wearing eyeglasses affects facial recognition', Current Psychology, 12(2), pp. 151–162. doi: 10.1007/BF02686820.

Tottenham, N. et al. (2009) 'The NimStim set of facial expressions: Judgments from untrained research participants', Psychiatry Research, 168(3), pp. 242–249. doi: 10.1016/j.psychres.2008.05.006.

Towler, A., White, D. and Kemp, R. I. (2017) 'Evaluating the feature comparison strategy for forensic face identification.', Journal of Experimental Psychology: Applied, 23(1), pp. 47–58. doi: 10.1037/xap0000108.

Urbanová, P. (2016) 'Performance of distance-based matching algorithms in 3D facial identification', Egyptian Journal of Forensic Sciences, 6(2), pp. 135–151. doi: 10.1016/j.ejfs.2016.04.004.

Valentine, T. (2006) 'Forensic facial identification', in Heaton-Armstrong, A. et al. (eds) Witness Testimony: Psychological, Investigative and Evidential Perspectives. Oxford, UK: Oxford University Press.

Valentine, T. and Davis, J. P. (2015) Forensic facial identification: theory and practice of identification from eyewitnesses, composites and CCTV, Forensic facial identification: theory and practice of identification from eyewitnesses, composites and CCTV. Edited by T. Valentine and J. P. Davis. Chichester, UK: John Wiley & Sons, Ltd. doi: 10.1002/9781118469538.

Vanezis, P. et al. (1996) 'Morphological Classification of Facial Features in Adult Caucasian Males Based on an Assessment of Photographs of 50 Subjects', Journal of Forensic Sciences, 41(5), p. 13998J. doi: 10.1520/jfs13998j.

Vanezis, P. and Brierley, C. (1996) 'Facial image comparison of crime suspects using video superimposition', Science and Justice - Journal of the Forensic Science Society, 36(1), pp. 27–33. doi: 10.1016/s1355-0306(96)72551-0.

Wall, M. (2019) 'Biased and wrong? Facial recognition tech in the dock', BBC News, 8 July, pp. 1–12. Available at: https://www.bbc.com/news/business-48842750. (Accessed on 16 March 2021)

Welsh, B. C. and Farrington, D. P. (2008) 'Effects of Closed-Circuit Television Surveillance on Crime', Campbell Systematic Reviews, 4(1), pp. 1–73. doi: 10.4073/csr.2008.17.

White, D., Dunn, J. D., et al. (2015) 'Error Rates in Users of Automatic Face Recognition Software', PLOS ONE. Edited by P. J. Hills, 10(10), p. e0139827. doi: 10.1371/journal.pone.0139827.

White, D., Jonathon Phillips, P., et al. (2015) 'Perceptual expertise in forensic facial image comparison', Proceedings of the Royal Society B: Biological Sciences, 282(1814). doi: 10.1098/rspb.2015.1292.

Wilkinson, C. and Evans, R. (2009) 'Are facial image analysis experts any better than the general public at identifying individuals from CCTV images?', Science & Justice, 49(3), pp. 191–196. doi: 10.1016/j.scijus.2008.10.011.

Wong, J. Y. et al. (2008) 'Validity and reliability of craniofacial anthropometric measurement of 3D digital photogrammetric images', Cleft Palate-Craniofacial Journal, 45(3), pp. 232–239. doi: 10.1597/06-175.

Yoshino, M. (2005) 'Facial image identification system based on 3D physiognomic data', in Clement, J. G. and Marks, M. K. (eds) Computer-graphic facial reconstruction. Boston: Elsevier Academic Press, pp. 347–362.

2. Development of the Wits Face Database: an African database of high-resolution facial photographs and multimodal closed-circuit television (CCTV) recordings

Nicholas Bacci, Joshua G. Davimes, Maryna Steyn, Nanette Briers

F1000Research, 2021, 10:131, https://doi.org/10.12688/f1000research.50887.1

F1000Research F1000Research 2021, 10:131 Last updated: 18 MAR 2021

Check for updates

DATA NOTE

Development of the Wits Face Database: an African database of high-resolution facial photographs and multimodal closed-circuit television (CCTV) recordings [version 1; peer review: 1 approved]

Nicholas Bacci, Joshua Davimes, Maryna Steyn, Nanette Briers

Human Variation and Identification Research Unit (HVIRU), School of Anatomical Sciences, University of the Witwatersrand, Johannesburg, Gauteng, 2193, South Africa

v1 **First published:** 19 Feb 2021, **10**:131
https://doi.org/10.12688/f1000research.50887.1
Latest published: 19 Feb 2021, **10**:131
https://doi.org/10.12688/f1000research.50887.1

Abstract

Forensic facial comparison is a commonly used, yet under-evaluated method employed in medicolegal contexts across the world. Testing the accuracy and reliability of facial comparisons requires large scale controlled and matching facial image databases. Databases that contain images of individuals on closed-circuit television (CCTV), with matching formal and informal photographs are needed for this type of research. Although many databases are available, the majority if not all are developed in order to improve facial recognition and face detection algorithms through machine learning, with very limited if any measure of standardisation. This paper aims to review the available databases and describe the development of a high resolution, standardised facial photograph and CCTV recording database of male Africans. The database is composed of a total of 6220 standardised and uncontrolled suboptimal facial photographs of 622 matching individuals in five different views, as well as corresponding CCTV footage of 334 individuals recorded under different realistic conditions. A detailed description of the composition and acquisition process of the database as well as its subdivisions and possible uses are provided. The challenges and limitations of developing this database are also highlighted, particularly with regard to obtaining CCTV video recordings and ethics for a database of faces. The application process to access the database is also briefly described.

Keywords

face database, CCTV, facial photographs, facial identification, facial comparison, morphological analysis, facial recognition

Open Peer Review

Reviewer Status ✔

	Invited Reviewers
	1
version 1 19 Feb 2021	✔ report

1. **Ching-Yiu Jessica Liu**, Liverpool John Moores University, Liverpool, UK

Any reports and responses or comments on the article can be found at the end of the article.

F1000Research F1000Research 2021, 10:131 Last updated: 18 MAR 2021

Corresponding author: Nicholas Bacci (nicholasbacci@gmail.com)

Author roles: Bacci N: Conceptualization, Data Curation, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Davimes J**: Data Curation, Investigation, Methodology, Project Administration, Resources, Visualization, Writing – Review & Editing; **Steyn M**: Conceptualization, Funding Acquisition, Methodology, Supervision, Writing – Review & Editing; **Briers N**: Conceptualization, Funding Acquisition, Methodology, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The research of N. Bacci (Grant No.: 11858) and N. Briers (Grant No.: CSUR160425163022 (UID:106031)) is sponsored by the National Research Foundation of South Africa. Any opinions, findings and conclusions or recommendations expressed in this study are those of the authors and therefore the NRF does not accept any liability in regard thereto. N. Bacci was also partially funded by the J.J.J. Smieszek Fellowship, School of Anatomical Sciences, University of the Witwatersrand.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 Bacci N *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

How to cite this article: Bacci N, Davimes J, Steyn M and Briers N. **Development of the Wits Face Database: an African database of high-resolution facial photographs and multimodal closed-circuit television (CCTV) recordings [version 1; peer review: 1 approved]** F1000Research 2021, **10**:131 https://doi.org/10.12688/f1000research.50887.1

First published: 19 Feb 2021, **10**:131 https://doi.org/10.12688/f1000research.50887.1

F1000Research 2021, 10:131 Last updated: 18 MAR 2021

Introduction

Facial comparison is utilised by law enforcement to associate two sets of images, captured on video or photographically, to one another. Although different approaches exist, facial comparison by morphological analysis is currently considered the most reliable method[1]. In this method, a target image, such as a snapshot from closed circuit television (CCTV) recordings obtained during criminal activity, and a standardised optimal image, such as a police mugshot, are compared to ascertain whether the two individuals are the same person. Facial comparison by morphological analysis has no directed standardised stepwise procedure as a validated methodology[2]. Although facial comparison still does employ the analysis, comparison, evaluation and verification (ACE-V) tenets of forensic examinations[1,3], its accuracy and reliability has not been tested extensively. As such, assessing its reliability should be considered a priority[4]. The lack of validation can potentially be attributed to the logistic complexity required for rigorous scientific testing, of which a considerable limitation is the lack of standardised and actualistic databases to use[5]. While several facial image databases exist (e.g. 5,6), the incongruity of their composition is a major limiting factor.

The composition of many face image databases tends to be specific to the original intended use with a focus on various controlled conditions which makes them difficult to use for general purposes[6]. For example, there is a variety of pre-landmarked databases available for use in the field of facial recognition (e.g. 7–11) with many variations and controlled conditions. Some of the most commonly controlled-for conditions are orientation of the head/pose, illumination/lighting conditions, facial expressions, and age-related variations (e.g. 6,7,12). Despite the large number of databases available (Table 1), there is a tendency towards either highly controlled data sets captured under very specific conditions with limited actualistic applications (e.g. 13,14) or highly randomised images collected under inconsistent conditions (e.g. 7–9,15). In addition, most of these databases include a limited number of subjects with many replications under niche conditions and no standard baseline control images. Some of the datasets with a limited number of unique individuals include under 100 subjects (e.g. 13,16,17), with a handful including only 10 to 15 subjects (e.g. 14,18–20). Lastly, many of these databases are, by today's standards, of subpar resolution with only two databases including images of resolutions greater than 640 x 480 pixels[21,22]. Between the highly specialised and varied conditions of capture, the lack of controlled images with matching realistic informal photographs of the same subjects, the low resolution, lack of methodological standardisation in image capture, and limited subject numbers, these facial image datasets provide very limited use in a forensic facial comparison context or more generalised facial studies.

Recently, databases consisting of video recordings of faces have become more common (e.g. 21,23–26) due to the increase in CCTV surveillance cameras. Generally, these databases have either been collated *via* recording of known participants in controlled environments (e.g. 21,23,26), or were based on pre-recorded videos obtained from various media on the internet or movies[24,25]. To the best of the authors' knowledge, only a single other database (SCface – Surveillance Cameras Face Database) published includes both facial photographs of high resolution and corresponding still images extracted from CCTV recordings of varying resolutions[21]. These databases are primarily used in testing and developing head/face detection and tracking or automated face annotation systems under a variety of video recording conditions.

The majority of facial image databases are primarily or exclusively inclusive of males of either European or East Asian descent (e.g. 17,27,28). Only a few databases contain individuals from other ancestry groups, particularly African individuals (e.g. 6,7,13), and they have a limited number of individuals. This is evident when considering there is only a single other South African face database[29]. Although a great initiative for a large, landmarked face database specifically developed to increase the variety of lighting, ethnicity and age available, it consists primarily of low-resolution webcam-based images with often very distorted lighting conditions. In addition, a total of 276 subjects were used for this database with no control images or demographic specifications provided.

The majority of existing facial databases have been developed for the purposes of facial recognition and machine learning training and do not contain target and control images of the same individual. These databases tend to be sourced from public, internet images containing faces in a variety of inconsistent conditions (e.g. 9,11,12). To the best of the authors' knowledge, no databases exist that are intended specifically for use with facial comparison by morphological analysis. As such, we aimed at creating a system for developing a consistent face database with corresponding individuals across various photographic and video recording conditions, resulting in the Wits Face Database (WFD). This database is intended to be a functional, actualistic African database of facial images that can be utilised for facial comparison analyses and research in craniofacial identification. This database is intended to be a free resource strictly for non-commercial scientific research, provided access has been cleared by an ethics committee overseeing its use.

Materials and methods

The database was established by collecting photographs of willing participants on the University of the Witwatersrand campus, Johannesburg, South Africa, including corresponding CCTV recordings. The database is comprised of CCTV recordings and photographs gathered on university premises, between July 2018 and October 2019, *via* the pre-existing CCTV systems used by campus security. Facial photographs were standardised and in five different views. The CCTV recordings were captured in a variety of conditions, such as from different quality cameras, in different formats and heights of recording, and with disguises (sunglasses and caps).

F1000Research 2021, 10:131 Last updated: 18 MAR 2021

Table 1. Overview of available face databases with available descriptives.

Database name and reference	No. of unique individuals	No. of images	Image resolution (pixels)	Database description and condition variations
AR Face Database[30]	126 (70 Male, 56 Female)	4000	576 x 768	Various facial expressions, lighting, glasses, scarf
CVL Database[31]	114	798	640 x 480	Various poses, varying facial expression
FERET Database[32]	1199	14051	256 x 384	Various slight facial expressions, poses
Labelled Faces in the Wild (LFW) Database[7]	5749	13233	250 x 250	Landmarked faces in various poses, expressions, lighting, ethnicity, age, clothing, hairstyles
Face Recognition Grand Challenge (FRGC) Database[33]	688	Undefined	"High resolution"	Various facial expression, lighting
CAS-PEAL Face Database[34]	1040 (595Male, 445 Female)	30900	360 x 480	Various poses, facial expressions, changes in lighting, glasses, caps
The MUCT Face Database[29]	276	3755	640 x 480	Variations in pose, lighting, and annotated faces
The Yale Face Database[18]	15	165	320 x 243	Facial expression variations, glasses
The Yale Face Database B[19]	10	5760	640 x 480	Various poses and changes in lighting
CMU Pose, Illumination, and Expression PIE Database[10]	68	41368	640 x 486	Various poses, facial expressions, changes in lighting, and glasses
Olivetti – Att – ORL[35]	40	400	92 x 112	None
Japanese Female Facial Expression (JAFFE) Database[20]	10	70	256 x 256	Various facial expressions
FIDENTIS 3D Face Database[36]	2476	2476 complete 3D scans	12 megapixels	3D scans, landmarked, ear to ear facial pose equivalent
Caltech Occluded Face in the Wild (COFW)[12]	Undisclosed	1852	Undisclosed	Various poses, expressions, lighting, occlusion focus, annotated
Ibug 300 Faces In-the-Wild (ibug 300W) Challenge database[9]	600	>4000	Undisclosed	Various poses, expressions, lighting, annotated
Labeled Face Parts in the Wild (LFPW) Dataset[11]	3000	3000	Undisclosed	Various poses, expressions, lighting
Quality labeled faces in the wild (QLFW) database[8]	5749	277809	250 x 250	Various poses, expressions, lighting, ethnicity, age, clothing, hairstyles, distortions
Helen dataset[37]	Undisclosed	2330	>500 width	Various poses, expressions, lighting, annotated
Facial expressions of emotion (KDEF) database[17]	70 (35 Male, 35 Female)	4900	Undisclosed	Various facial expressions
NimStim facial expression database[13]	43	672	Undisclosed	Various facial expressions

Database name and reference	No. of unique individuals	No. of images	Image resolution (pixels)	Database description and condition variations
Annotated Facial Landmarks in the Wild (AFLW) database[15]	25993 (11437 Male, 14556 Female)	21997	Undisclosed	Various poses, expressions, lighting, ethnicity, age, clothing, and hairstyles
Pointing Head Pose Image Database[14]	15	2790	384 x 288	Various poses, glasses
BioID Database[38]	23	1521	382 x 288	Various facial expressions, poses, lighting, accessories
University of Olulu Physics-Based Face Database[39]	111	2112	428 x 569	Various poses (minor), lighting, glasses
Chicago Face Database[22]	158 (73 Male, 85 Female)	Undisclosed	2444 x 1718	Various poses, facial expressions, lighting
SCface – Surveillance Cameras Face Database[21]	130 (114 Male, 16 Female)	4160	Varied from 3072 x 2048 to 224 x 168	Various poses, normal and infrared
M2VTS multimodal face database[16]	37	Minimum of 185	286 x 350	Videos and video to still images, various poses, some with and without eyeglasses, various time intervals (weeks)

F1000Research 2021, 10:131 Last updated: 18 MAR 2021

Ethics and participant recruitment

Ethical approval was obtained from the Human Research Ethics Committee (Medical) of the University of the Witwatersrand (clearance certificate No.: M171026). Permits from the campus head of security and deputy registrar were obtained prior to data capture, in their capacity as site managers. Facial photographs were captured by an experienced photographer following recruitment of participants. CCTV recordings of matched participants were collected through the university's security systems. The original target of participants intended for the database were 600. However, due to data loss as a result of power failures and data transfer corruption a greater number of participants had to be recruited for redundancy purposes. Potential participants were identified and recruited near the data collection sites based on facial anthropological features resembling males of South African descent and being older than 18 years of age. Approached participants were then informed of the greater project orally and with an information sheet[40] and asked for voluntary participation in the study and database. Once agreed, they signed an informed consent form[40] prior to being photographed or recorded. If participants requested to be removed from the database, all their photographs and recordings were erased and their recruitment information shredded. None of the personal details and images given by the participants are to be freely distributed. As agreed with the ethics committee, and according to the consent forms signed by the participants, no identifiable images of any individuals are to be published without following up with participants for additional consent. As a result, for publication purposes to indicate examples of the face database, one of the author's images have been utilised.

Image acquisition

Facial photographs and recordings were collected at three designated access-controlled locations with a large influx of potential participants on the Braamfontein Campus of the University of the Witwatersrand, Johannesburg, South Africa. The following three collection sites were set-up and utilised:

- Site A: outdoors CCTV camera (installation height: 3100 mm) with a view of a student card terminal near the food concourse and Student Union Building, the Matrix.
- Site B: outdoors CCTV camera positioned at eye-level height (1700 mm) in one of the more frequented pedestrian entrances in proximity to the Oppenheimer Life Sciences building.
- Site C: indoors analogue CCTV camera (installation height: 2500 mm) in the administration building concourse in view of the cashier's offices, Solomon Mahlangu House.

In the vicinity of the installed CCTV cameras at each site, a photography station was set up in a standard manner as demonstrated in Figure 1 and Figure 2. Within this set-up, facial photographs were captured in five different views using two different cameras with slightly different parameters and conditions. All cameras were arranged in a marked fixed location near the CCTV field of view on tripods at a fixed height of 1600 mm. This height was maintained to attempt centring the field of view of the photographs on the face, as the mean height for black South African males[41] is 1710 mm. Specifically, the eye level of each participant was composed on the top horizontal rule of thirds line for all photographs taken. All photographs were taken in portrait orientation.

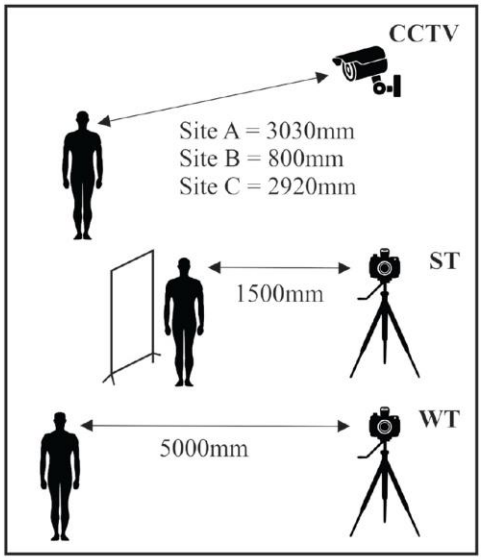


Figure 1. Schematic diagram of camera set-up for closed-circuit television (CCTV) and facial photograph capture. ST = standardized photograph; WT = wildtype photograph.

An example of the arrangements for database image capture and recording on pre-existing CCTV systems is shown in Figure 2. Distances were controlled for participants to be photographed and recorded at each site. Standardised (ST) photographs, with a solid black backdrop and participant clothing covered, were captured at a distance of 1500 mm. Wildtype (WT) photographs were captured at a 5000 mm distance. These photographs included a mixed background that was intentionally meant to simulate real-life photographic conditions. WT photographs were taken in a simulated scenario of suboptimal facial images with a comparable quality camera and facial poses. The background was purposefully not controlled for with a mixed environment visible and varied based on the location site of data collection. This background was intentionally out of focus to simulate suboptimal photographic conditions. Despite a minor level of variation, consistency was maintained across all photographs with regard to distance to subject, aperture, and composition.

The first set of photographs were captured under standardised conditions using a Canon 1300D 18MP DSLR camera (18–55 mm DC Canon lens) with the following settings: image

F1000Research 2021, 10:131 Last updated: 18 MAR 2021



Figure 2. Actual photographic camera set-up in the process of database development. Arrangement of cameras and backdrop for standardised and wildtype photographs at site A (**a**), site B (**b**) and site C (**c**).

sensor sensitivity (ISO) of 800, aperture F/9, shutter speed between 1/125 and 1/40, focal length of 55 mm and daylight white balance. For these standard photographs, the objective to face distance was fixed at 1500 mm. A set of standardised photographs were taken, with a black backdrop and with the participants' clothing covered by a black velvet cloth – similar to the backdrop – in order to prevent matching participants based on clothing appearance. These standard photographs were captured in the following five views (Figure 3):

a. Anterior frontal view

b. Right 45-degree view

c. Right lateral view

d. Left 45-degree view

e. Left lateral view

The second set of WT photographs captured for each participant, corresponded to the same views as described above, using a Sony SLT A57 (18 – 250 mm Sony Zoom Lens) with the following settings: image sensor sensitivity (ISO) of 200, aperture F/9, shutter speed between 1/125 and 1/40, focal length of 250 mm and daylight white balance. Indoor photographs at site C were captured at a range of ISOs between 400 and 1600 depending on the varying light conditions.

All photographs were captured as both native .jpeg and RAW format. Images were then downloaded from the SD cards of each camera, stored, and sorted on a desktop computer. All images were then batch cropped at a 4x5 (8x10) aspect ratio, using Adobe Photoshop CS6, to only include the participants as centred in the frame of the photograph. The resulting image resolution for the ST photographs was 3456 x 4320 pixels at 300 dpi and sRGB colour representation. The WT

F1000Research 2021, 10:131 Last updated: 18 MAR 2021



Figure 3. Example of the five views of standardised (left) and wildtype (right) facial photographs captured. The five views of facial photographs are demonstrated in this image, showing anterior (**a**), right 45-degree (**b**), right lateral (**c**), left 45-degree (**d**) and left lateral (**e**) views.

photographs' resolution was 3264 x 4080 at 350 dpi and with sRGB colour representation. Following batch cropping, the standardised images were imported into Adobe Lightroom (v. 5.3) for basic editing. The only adjustments made included exposure level, highlights, and shadow correction. Additionally, removal/spot healing of any exposed clothing or background features were done if the cloth or backdrop did not fully obscure the participant and surroundings. The wildtype images were left unaltered post cropping. The above discussed image processing (batch cropping) and adjustments (exposure level, highlights, shadows correction and spot healing) can alternatively be performed using open source software such as GNU Image Manipulation Program (GIMP), Photivo, and darktable.

CCTV recordings from internet protocol (IP) cameras were transferred live to a HikVision (model: DS-9664NI-I16) server for storage. Videos from the analogue camera were stored on a DS-ENC-V120B20121026-7054D2ABA6FB digital video

F1000Research 2021, 10:131 Last updated: 18 MAR 2021

recorder (DVR) device. All recordings were then extracted in .mp4 format from the University of the Witwatersrand's CCTV systems through the university's protection services software (iVMS 4200 v. 2.71.9). Only footage recorded during data collection times was extracted from the CCTV systems. During recording, participants were asked to stand at a marked location and rotate through the same five views that the photographs were captured in for approximately 5 seconds. Following this, the participant acted out the process of utilising a student card terminal in view of the camera (sites A and C) or attempting to exit the campus (site B).

The outdoor CCTV camera at site A was an IP camera (HikVision, model: DS-2CD2142FWD-I, 4-megapixel, 4 mm fixed lens, aperture F/2). This camera was installed at a height of 3100 mm and a floor distance of 2690 mm from the marked location where participants stood in the process of video capturing. The resulting distance between the participants' face and the IP camera objective was 3030 mm (Figure 4). This distance was calculated based on the mean height of the average black South African male of 1710 mm[41]. The angle of incidence from the camera objective to the face was approximately 27°.

The outdoor "eye-level" IP CCTV camera at site B had the same model and specifications of the camera at site A and was installed at a height of 1700 mm from the ground, marked location to floor distance of 800 mm. The resulting distance from target face to camera was 800 mm and the angle of incidence was virtually zero.

The indoor analogue CCTV camera at site C (Securi-Prod 1/3" Sony Effio E 700TVL indoor dome, model: CC217, 2.8 – 12 mm vari-focal lens) was installed at a height of 2500 mm and at a floor distance from the marked area of 2810 mm. The resulting distance between camera and face was 2920 mm and the angle of incidence was approximately 22°.

A total of approximately 30 seconds of footage was recorded for each participant and out of the 30 seconds five still images were captured from the footage at each of the five views previously described (Figure 5, Figure 6, Figure 7 and Figure 8). The photographs and recordings of each individual were



Figure 4. Demonstration of estimated objective to face distance at site A. Vertical distance indicates the mean height of a South African male (1710 mm) and the oblique distance indicates the calculated approxiamte camera to face distance at Site A (3030 mm).



Figure 5. Example of the five views of standard closed-circuit television (CCTV) stills captured at site A. The five views of facial photographs are demonstrated in this image, showing anterior (**a**), right 45-degree (**b**), right lateral (**c**), left 45-degree (**d**) and left lateral (**e**) views.

F1000Research 2021, 10:131 Last updated: 18 MAR 2021

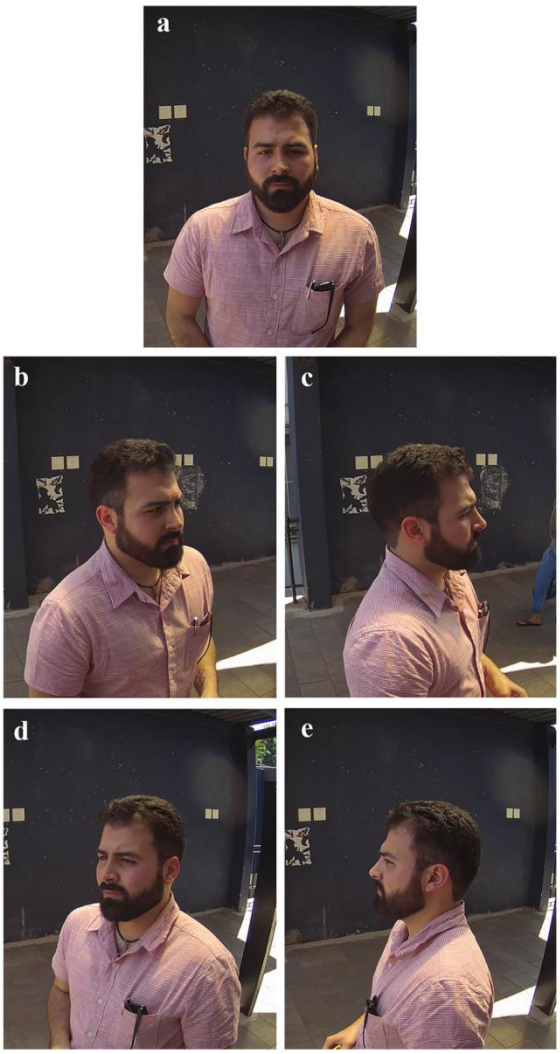


Figure 6. Example of the five views of eye level closed-circuit television (CCTV) stills captured at site B. The five views of facial photographs are demonstrated in this image, showcasing anterior (**a**), right 45-degree (**b**), right lateral (**c**), left 45-degree (**d**) and left lateral (**e**) views.



Figure 7. Example of the five views of analogue closed-circuit television (CCTV) stills captured at site C. The five views of facial photographs are demonstrated in this image, showing anterior (**a**), right 45-degree (**b**), right lateral (**c**), left 45-degree (**d**) and left lateral (**e**) views.

coded with a participant number to maintain anonymity of the participants and a separate record of the identity of the participants was retained. Complete anonymity of the participants included in the facial image database was maintained. Images and videos included in the database were stored on three separate password protected desktop computers and encrypted on a constantly monitored external hard disk drive. In addition, the dataset was also transferred to the University of the Witwatersrand's Library repository under restricted access to data management services. This repository retains the database in triplicate for cataloguing and in order to allocate searchable metadata to it to facilitate use.

Database composition

The database was composed strictly of male South Africans of African descent. Participants were over the age of 18, for consent purposes, and in the young adult age-ranges, between 18 and 35 years of age. The participants were subdivided

F1000Research 2021, 10:131 Last updated: 18 MAR 2021

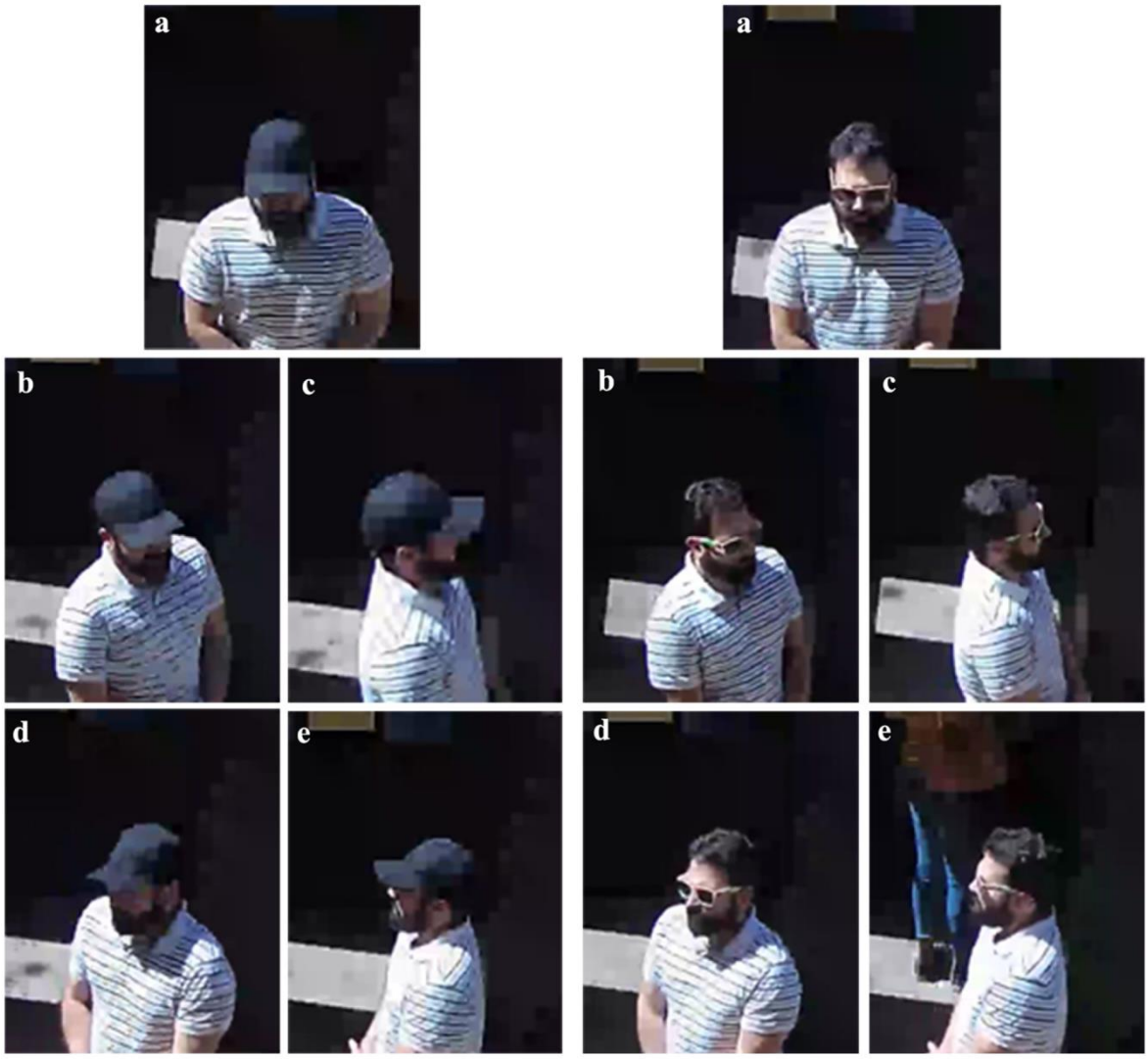


Figure 8. Example of the five views of closed-circuit television (CCTV) recordings with two obstruction types - brimmed cap (left) and sunglasses (right). The five views of facial photographs are demonstrated in this image, showing anterior (**a**), right 45-degree (**b**), right lateral (**c**), left 45-degree (**d**) and left lateral (**e**) views.

into a series of cohorts depending on the type of matched analysis possible, as outlined in Table 2. All participants were photographed in both the standardised and wildtype setting at the various sites. The first two groups were only photographed under natural outdoor lighting at site A (n=120) as well as with artificial indoor fluorescent lighting at site C (n=99). For each participant 10 photographs were captured, totalling 1200 photographs under outdoor conditions with natural lighting conditions and 990 photographs under indoor fluorescent lighting conditions. A second group, in addition to being photographed as above, was recorded on an outdoors setting CCTV camera at site A (n=98, n=86 with corresponding footage). A third group was recorded at site B with the eye-level camera (n=108, n=76 with corresponding footage). A fourth cohort was recorded at site C with the analogue CCTV camera and lastly a final group of participants were recorded with obstructive accessories, namely caps (n=45, n=34 with corresponding footage) or sunglasses (n=41, n=31 with corresponding footage),

F1000Research 2021, 10:131 Last updated: 18 MAR 2021

Table 2. Detailed categorisation of Wits Face Database (WFD) composition by cohorts of potential analyses.

Cohort of matching individuals	No. of participants	No. of photographs	Participants with corresponding CCTV footage
Photo-Photo Outdoor (Site A)	120	1200	0
Photo-Photo Indoor (Site C)	99	990	0
Photo-Photo Totals	**219**	**2190**	**0**
Photo-CCTV Outdoor (Site A)	98	980	89
Photo-CCTV Outdoor Eye Level (Site B)	108	1080	76
Photo-CCTV Indoor Analogue (Site C)	111	1110	107
Photo-CCTV Outdoor with Cap (Site A)	45	520	34
Photo-CCTV Outdoor with Sunglasses (Site A)	41	490	31
Photo-CCTV Totals	**403**	**4030**	**334**
Grand Totals	**622**	**6220**	**334**

CCTV= closed-circuit television.

using the same IP camera at site A. Due to data corruption or data loss in the CCTV recording process, not all photographed individuals have corresponding footage associated to them. Overall, the database is inclusive of over 6200 facial photographs and 334 corresponding video recordings from various types of CCTV cameras within an African sample (Table 2).

Database utility and applications

A database of this scale can be utilised in a variety of training and research applications, particularly when considering this as the first database of facial images with such a large complement of facial photographs of African individuals. Among a variety of possible applications, the primary intended use is testing various methodologies and conditions for forensic facial comparison. Furthermore, this database can be utilised as a prime tool for training facial comparison experts to develop a high rate of competency and proficiency. Having an appropriate level of training is a crucial aspect of the judicial process[1,4]. In addition, the faces in this database could be used to generate stimuli for studies in psychology and marketing sciences relating to facial recognition. Likewise, the images in this database could be modified as required on an *ad hoc* basis and implemented to train and develop future machine learning and artificial intelligence systems for the purpose of facial recognition in a forensic context. Although a potentially laborious process, standardised landmarks could be added to allow for an estimated dimension calculation of the various facial proportions and features as required by facial recognition systems. However, the database is only available for *bona fide* researchers affiliated with academic research institutions and not for commercial use. An example use the dataset was designed for was validating the use of morphological analysis in forensic facial comparison across a photographic and CCTV sample[27]. This was achieved by sub-setting selected photographs and CCTV recording stills from the database into facial image pools that were independently analysed[27] following the Facial Identification Scientific Working Group morphological analysis feature list[42].

Challenges and limitations

Developing a database of facial photographs and surveillance footage is a logistically complex and tedious process. It can be difficult to recruit volunteers even in highly trafficked areas in a major university (the University of the Witwatersrand has over 39,500 students across five campuses), although people on a campus tend to feel more secure and therefore more willing to participate in this type of study. The recruitment and photography process require a large amount of manpower with a minimum of three to four volunteers needed per day as assistants for an efficient image acquisition process. The location management and site selection are also a highly demanding task as one needs to limit variations between recorded images as much as possible. This entire process needs to be carried out while still collecting images that are representative of a somewhat realistic scenario. The variations in conditions of recordings vary based on camera type, quality, and installation. The majority of CCTV recordings, for example, were collected in an outdoors area with sunny daylight conditions which is by design an actualistic sample, although inconsistent due to weather and lighting conditions changing throughout a given day.

Similarly, the photography collection was affected by uncontrollable varying lighting conditions for both the indoor and outdoor settings. These include dim lighting indoors leading

F1000Research 2021, 10:131 Last updated: 18 MAR 2021

to lower quality, higher noise images, and dappled light and sun position outdoors leading to highlights, shadow, and contrast artefacts. The objective to face distances were controlled as much as possible, without compromising the actualistic nature of the data, in order to minimise perspective distortion due to the nature of relaying a three-dimensional scene/object into a two-dimensional medium of photography[43]. Despite the intent, at greater distances the focal plane of a camera lens requires adjusting and can result in lower clarity images[43], as is evident in the varied focal lengths of the CCTV cameras.

Furthermore, the nature of CCTV cameras depends on the entirety of the surveillance system installation. A variety of complications, including inconsistent IP camera network connectivity and power outages caused occasional complete data loss or footage corruption in the recording process, resulting in the overall reduced numbers of corresponding recordings. This data loss was particularly evident in the IP cameras, as in the particular set-up used by the university, they do not locally store footage but transfer it immediately to a central server. In the process, any interruption or fluctuation in local area network speeds or connectivity would result in data loss or corruption. Even though analogue cameras by default record at lower resolutions, the immediate local storage on a DVR device resulted in reduced data loss and no data corruption. In fact, a total of 710 participants were originally recruited and photographed and/or recorded. Following data loss, participants that requested to be excluded from the database and data corrupted during transfer from the CCTV cameras to the servers, only 622 of the 710 could be included in the database.

Male individuals were selected specifically as males are more commonly involved in criminal activity, both as victims and perpetrators, in both developed and developing nations[44,45]. However, other demographic factors such as age and ancestry were not strictly controlled, and sample composition resulted varied across all cohorts. This lack of strict control was primarily due to the sensitivity of labelling groups of individuals as belonging to specific descent groups, making requesting this information from participants one of the ethical limitations considered originally.

The intention of the authors is for this database to be further expanded to include female individuals, additional disguises in the form of make-up and face masks, as well as more variations of CCTV camera recordings, such as infrared night vision CCTV, to provide a better and more varied selection for its applications.

Ethics of face databases

Overall, face databases are quite common with 27 photographic image databases available for use in the fields of head and face detection, tracking and recognition. These databases are usually created with specific criteria, such as particular facial expressions and lighting condition variations for testing a specific aspect of facial recognition. Amongst the 27 databases outlined in Table 1, seven have been collected from various media of personal photographs available on the internet[7,9,11,12,15,24,37], which is a legal yet perhaps ethically questionable practice, as the individuals included in these face databases have not provided consent for that inclusion. This is particularly common in the most recently developed databases (e.g. 7,8,15), most likely due to the broad availability of facial images and the highly efficient and accurate search engines. This practice can result in privacy-intrusive practices of freely available and commercialised application of facial recognition[28], which is an ever-increasing concern for the public with specific regard to privacy of data and images[46]. The database compiled here is an example of a complex yet more ethical approach that limits commercialised application and potential research.

Data availability

Underlying data

The WFD is stored on the Wits Institutional Repository environment on DSpace (WIReDSpace) and published under the following unique identifier: http://doi.org/10.17605/OSF.IO/WMA4C (this registration also contains the PhD study protocol that led to the development of the WFD and an addendum to the protocol registration highlighting the major changes to the methodological approach of the original protocol). A sample of the dataset is freely accessible at https://hdl.handle.net/10539/29924.

The database is an open access resource for use in strictly non-commercial research. In order to access the WFD, prospective users will have to apply for access to the Institutional Review Board overseeing ethical and scientific use of the database in order to safeguard the privacy and decency of the database's participants. Once approved a researcher may use the database free of charge. Database access is restricted and limited to following the above-mentioned procedure, due to the nature of the data including potentially identifying information (facial physiognomy) of participants. In addition, strict limitations were imposed by the Human Research Ethics Committee (Medical) as well as the consent permissions agreed upon with the participants, which assign responsibility to the School of Anatomical Sciences to review access applications in ethical and scientific merit in order to exclusively conduct research. The access procedures and limitations are governed by a legally binding Conditions of Use document available on https://hdl.handle.net/10539/29924[40] in conjunction to the freely accessible sample. Data will be made available to successful applicants under a temporary restricted licence guided by the aforementioned conditions of use document.

Extended data

Open Science Framework: Wits Face Database: Description. https://doi.org/10.17605/OSF.IO/Q8V2R[40]

This project contains the following extended data:

- Participant information sheet
- Participant consent form
- Conditions of use

F1000Research 2021, 10:131 Last updated: 18 MAR 2021

Data are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

Authors' contributions

Conceptualization: Nicholas Bacci, Maryna Steyn, Nanette Briers; Data curation: Nicholas Bacci, Joshua Davimes; Methodology: Nicholas Bacci, Joshua Davimes, Maryna Steyn, Nanette Briers; Investigation: Nicholas Bacci, Joshua Davimes; Project Administration: Nicholas Bacci, Joshua Davimes; Resources: Nicholas Bacci, Joshua Davimes; Visualisation: Nicholas Bacci, Joshua Davimes; Writing - original draft preparation: Nicholas Bacci; Writing - review and editing: Nicholas Bacci, Joshua Davimes, Maryna Steyn, Nanette Briers; Funding acquisition: Nicholas Bacci, Maryna Steyn, Nanette Briers; Supervision: Maryna Steyn and Nanette Briers.

Acknowledgements

Thanks are due to all the participants who agreed to be photographed for the development of this database. Particular recognition is due to all the volunteers who assisted in participant recruitment for this study: Jesse Fredericks, Kiveshen Pillay, Rethabile Masiu, Sameerah Sallie, Daniel Munesamy, Laurette Joubert, Jordan Swiegers, Betty Mkabela, Johannes P. Meyer, Amy Spies, Natasha Loubser, Nicole Virgili, Dan-Joel Lukumbi, Tamara Lottering, Mathabatha Ntjie, Claudia Landsman, Raheema Dalika, Merete Goosen, Stephanie Souris, Rabelani Negota, Mahlatse Mahasha, Jessica Manavhela. Special thanks are due to Gideon LeRoux for his assistance in the capturing and extracting of the CCTV recordings from the University security systems. A great deal of thanks are also due to Nina Lewin who aided in policy development and establishing the database repository. All named persons above have granted permission to be included within the acknowledgements.

References

1. Facial Identification Scientific Working Group: **Facial Comparison Overview and Methodology Guidelines.** 2019.
Reference Source
2. Dror IE, Charlton D, Péron AE: **Contextual information renders experts vulnerable to making erroneous identifications.** *Forensic Sci Int.* 2006; **156**(1): 74–8.
PubMed Abstract | Publisher Full Text
3. Speckeis C: **Can ACE-V be validated?** *J Forensic Identif.* 2011; **61**(3): 201–9.
Reference Source
4. Steyn M, Pretorius M, Briers N, *et al.*: **Forensic facial comparison in South Africa: State of the science.** *Forensic Sci Int.* 2018; **287**: 190–4.
PubMed Abstract | Publisher Full Text
5. Valentine T, Davis JP: **Forensic Facial Identification.** *Forensic Facial Identification.* 2015; 1–347.
6. Gross R: **Face Databases.** *Handbook of Face Recognition*. New York: Springer-Verlag; 2005; 301–27.
Publisher Full Text
7. Huang GB, Ramesh M, Berg T, *et al.*: **Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments.** October, University of Massachusetts, Amherst Technical Report. 2007.
Reference Source
8. Karam LJ, Zhu T: **Quality labeled faces in the wild (QLFW): a database for studying face recognition in real-world environments.** *Hum Vis Electron Imaging XX.* 2015; **9394**: 93940B.
Publisher Full Text
9. Sagonas C, Antonakos E, Tzimiropoulos G, *et al.*: **300 Faces In-The-Wild Challenge: database and results.** *Image Vis Comput.* 2015; **47**: 3–18.
Publisher Full Text
10. Sim T, Baker S, Bsat M: **The CMU Pose, Illumination, and Expression (PIE) database.** In: *IEEE Trans Pattern Anal Mach Intell.* 2003; 1615–8.
Publisher Full Text
11. Belhumeur PN, Jacobs DW, Kriegman DJ, *et al.*: **Localizing parts of faces using a consensus of exemplars.** *IEEE Trans Pattern Anal Mach Intell.* 2013; **35**(12): 2930–40.
Publisher Full Text
12. Burgos-Artizzu XP, Perona P, Dollar P: **Robust face landmark estimation under occlusion.** *Proc IEEE Int Conf Comput Vis.* 2013; 1513–20.
Publisher Full Text
13. Tottenham N, Tanaka JW, Leon AC, *et al.*: **The NimStim set of facial expressions: Judgments from untrained research participants.** *Psychiatry Res.* 2009; **168**(3): 242–9.
PubMed Abstract | Publisher Full Text | Free Full Text
14. Gourier N, Hall D, Crowley JL: **Estimating face orientation from robust detection of salient facial structures.** *International Workshop on Visual Observation of Deicitic Gestures.* 2004; 17–25.
Reference Source
15. Kostinger M, Wohlhart P, Roth PM, *et al.*: **Annotated Facial Landmarks in the Wild: a large-scale, real-world database for facial landmark localization.** *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops).* IEEE; 2011; 2144–51.
Publisher Full Text
16. Pigeon S, Vandendorpe L: **The M2VTS Multimodal Face Database.** *First Int Conf Audio-and Video-based Biometric Pers Authentication.* 1997; 403–9.
17. Calvo MG, Lundqvist D: **Facial expressions of emotion (KDEF): Identification under different display-duration conditions.** *Behav Res Methods.* 2008; **40**(1): 109–15.
PubMed Abstract | Publisher Full Text
18. Belhumeur PN, Kriegman DJ: **Eigenfaces vs. Fisherfaces : Recognition Using Class Specific Linear Projection.** *IEEE Trans Pattern Anal Mach Intell.* 1997; 711–20.
Publisher Full Text
19. Georghiades AS, Belhumeur PN, Kriegman DJ: **From few to many: Illumination cone models for face recognition under variable lighting and pose.** *IEEE Trans Pattern Anal Mach Intell.* 2001; **23**(6): 643–60.
Publisher Full Text
20. Lyons M, Akamatsu S, Kamachi M, *et al.*: **Coding facial expressions with Gabor wavelets.** *Proc - 3rd IEEE Int Conf Autom Face Gesture Recognition, FG 1998.* 1998; 200–5.
Publisher Full Text
21. Grgic M, Delac K, Grgic S: **SCface - Surveillance cameras face database.** *Multimed Tools Appl.* 2011; **51**(3): 863–79.
Publisher Full Text
22. Ma DS, Correll J, Wittenbrink B: **The Chicago face database: A free stimulus set of faces and norming data.** *Behav Res Methods.* 2015; **47**(4): 1122–35.
PubMed Abstract | Publisher Full Text
23. La Cascia M, Sclaroff S, Athitsos V: **Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3D models.** In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on.* 2002; **22**(4): 322–36.
Publisher Full Text
24. Dhall A, Dhall A, Member S, *et al.*: **Collecting Large, Richly Annotated Facial-Expression Databases from Movies.** *J LaTeX Cl Files.* 2007; **6**(1): 1–14.
Reference Source
25. Shen J, Zafeiriou S, Chrysos GG, *et al.*: **The First Facial Landmark Tracking in-The-Wild Challenge: Benchmark and Results.** In: *IEEE International Conference on Computer Vision.* 2016; 1003–11.
Publisher Full Text
26. Ariz M, Bengoechea JJ, Villanueva A, *et al.*: **A novel 2D/3D database with automatic face annotation for head tracking and pose estimation.** *Comput*

Vis Image Underst. 2016; **148**: 201–10.
Publisher Full Text

27. Bacci N, Houlton TMR, Briers N, *et al.*: **Validation of forensic facial comparison by morphological analysis in photographic and CCTV samples.** *Int J Legal Med.* 2021. (In Press).

28. Senior AW, Pankanti S: **Privacy Protection and Face Recognition**. In: Huang T, Xiong Z, Zhang Z. editors. *Handbook of Face Recognition.* 2nd ed. London: Springer; 2011; 671–91.
Publisher Full Text

29. Milborrow S, Morkel J, Nicolls F: **The MUCT Landmarked Face Database**. Pattern Recognit Assoc South Africa. 2008.
Reference Source

30. Martinez AM, Benavente R: **The AR Face Database CVC Tech. Report #24**. 1998; 24.
Reference Source

31. Solina F, Peer P, Batagelj B, *et al.*: **Color-based face detection in the" 15 seconds of fame" art installation**. *Proc Mirage, Conf Comput Vision/Computer Graph Collab Model Imaging, Render Image Anal Graph Spec Eff Rocquencourt, Fr.* 2003; 38–47.
Reference Source

32. Phillips JP, Moon H, Rizvi SA, *et al.*: **The FERET evaluation methodology for face-recognition algorithms.** In: *IEEE Trans Pattern Anal Mach Intell.* 2000; **22**(10): 1090–1104.
Publisher Full Text

33. Phillips PJ, Flynn PJ, Scruggs T, *et al.*: **Preliminary face recognition grand challenge results.** In: *International Conference on Automatic Face and Gesture Recognition.* 2006; 15–21.
Publisher Full Text

34. Gao W, Cao B, Shan S, *et al.*: **The CAS-PEAL large-scale chinese face database and baseline evaluations.** *IEEE Trans Syst Man Cybern Syst.* 2008; **38**(1): 149–61.
Reference Source

35. Samaria FS, Harter AC: **Parameterisation of a stochastic model for human face identification.** *IEEE Work Appl Comput Vis - Proc.* 1994; 138–42.
Publisher Full Text

36. Urbanová P, Ferková Z, Jandová M, *et al.*: **Introducing the FIDENTIS 3D Face Database.** *Anthropol Rev.* 2018; **81**(2): 202–23.
Publisher Full Text

37. Le V, Brandt J, Lin Z, *et al.*: **Interactive Facial Feature Localization.** In: *European Conference on Computer Vision.* 2012; 679–92.
Publisher Full Text

38. Jesorsky O, Kirchberg KJ, Frischholz RW: **Robust face detection using the Hausdorff distance.** In: *International Conference on Audio- and Video-based Biometric Person Authentication.* 2001; 90–5.
Publisher Full Text

39. Marszalec E, Martinkauppi B, Soriano M, *et al.*: **Physics-based face database for color research.** *J Electron Imaging.* 2000; **9**(1): 32–28.
Publisher Full Text

40. Lewin NS, Bacci N, Davimes J, *et al.*: **Wits Face Database: Description**. 2021.
http://www.doi.org/10.17605/OSF.IO/Q8V2R

41. Steyn M, Smith JR: **Interpretation of ante-mortem stature estimates in South Africans.** *Forensic Sci Int.* 2007; **171**(2–3): 97–102.
PubMed Abstract | Publisher Full Text

42. Facial Identification Scientific Working Group: **Facial Image Comparison Feature List for Morphological Analysis**. 2018.
Reference Source

43. Stephan CN: **Perspective distortion in craniofacial superimposition: Logarithmic decay curves mapped mathematically and by practical experiment.** *Forensic Sci Int.* 2015; **257**: 520.e1–520.e8.
PubMed Abstract | Publisher Full Text

44. Cooper A, Smith E: **Homicide Trends in the United States,1980-2008: Annual Rates for 2009 and 2010**. U.S. Department of Justice, Office of Justice Programs, Beuraeu of Justice Statistics. 2011.
Reference Source

45. Maluleke R: **Crime Statistics Series Volume V: Crime against Women in South Africa**. 2018.
Reference Source

46. Finn RL, Wright D, Friedewald M: **Seven Types of Privacy**. *European Data Protection: Coming of Age.* 1st ed. Dordrecht: Springer Science+Business Media; 2013; 3–32.
Publisher Full Text

F1000Research F1000Research 2021, 10:131 Last updated: 18 MAR 2021

Open Peer Review

Current Peer Review Status: ✔

Version 1

Reviewer Report 05 March 2021

https://doi.org/10.5256/f1000research.53979.r79950

© **2021 Liu C.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Ching-Yiu Jessica Liu**
Liverpool John Moores University, Liverpool, UK

This article described the development of the 'Wits Face Database'. The authors collected CCTV footage from three different environments, along with standardised and uncontrolled facial photographs detailed in Table 2. A total of 622 adult male individuals with facial anthropological features resembling South African descent were recruited.

It will be beneficial to document the framerates of the CCTV recordings.

The mean height for black South African males was used for the photography station for ST images. Did you have to make any adjustments for different heights or were all participants within the capture range?

"individuals with facial anthropological features resembling South African descent were recruited." I am not sure I fully understand what this includes and excludes. You mentioned that age and ancestry were not strictly controlled along with the sensitivity of labelling groups. Does this mean anyone who was an adult male was recruited?

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Facial Anthropology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 17 Mar 2021

Nicholas Bacci, University of the Witwatersrand, Johannesburg, South Africa

Dear Dr Liu,

Thank you for your prompt and positive review. We appreciate your insights for the missing information and points requiring clarification. We will be including the following information in a new version of the manuscript to provide the information to all readers:

1. Comment: It will be beneficial to document the framerates of the CCTV recordings. Response: The framerates and codec information for the IP camera CCTV recordings will be included. However, the analogue camera did not store framerate or codec information in its metadata or the manufacturer website. Amendment: The following statements will be included in the revised manuscript version for clarity: “The recordings from these two cameras were encoded in MPEG-H Part2/HEVC (H.265) codec and had a frame rate of 20 frames per second.”. “No codec and framerate information was available for this camera both on the manufacturer website and the recording metadata.”
2. Comment: The mean height for black South African males was used for the photography station for ST images. Did you have to make any adjustments for different heights or were all participants within the capture range? Response: The height of the tripod for the ST and WT photographs was not adjusted, it was maintained fixed at the average height of black South African males throughout the data collection process. Only a handful of participants were significantly above or below the mean height in order to affect the composition of the photographs. In those cases the angle of the camera was adjusted slightly in the vertical axis to retain the same composition. Amendment: The following clarification will be included in the manuscript: “In all cases, for both ST and WT photographs, the tripod height was not adjusted and left at the fixed height of 1600mm. The majority of participants were within the capture range; however, for a few of the individuals who were either vastly below or above average South African male height, the camera was tilted in the vertical axis of the tripod in order to align their eyes along the uppermost 1/3rd of the portrait composition. No other changes were made in terms of focal length or post crop editing.”
3. Comment: "individuals with facial anthropological features resembling South African descent were recruited." I am not sure I fully understand what this includes and excludes. You mentioned that age and ancestry were not strictly controlled along with the sensitivity of labelling groups. Does this mean anyone who was an adult male was recruited? Response: Clarification on the participants included in the study will be included. Male South Africans of facial appearance resembling African descent were specifically recruited for the database. Amendment: The statement speaking to

participant descent will be amended for clarity as follows: "Potential participants were identified and recruited near the data collection sites based on facial anthropological features resembling South African males of African descent and being older than 18 years of age.".

Best Regards,
Nicholas Bacci

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research

3. Validation of forensic facial comparison by morphological analysis in photographic and CCTV samples

Nicholas Bacci, Tobias M. R. Houlton, Nanette Briers, Maryna Steyn

International Journal of Legal Medicine, 2021, https://doi.org/10.1007/s00414-021-02512-3

Validation of forensic facial comparison by morphological analysis in photographic and CCTV samples

Nicholas Bacci[1], Tobias MR Houlton[1], Nanette Briers[1], Maryna Steyn[1]

[1]Human Variation and Identification Research Unit, School of Anatomical Sciences, University of the Witwatersrand, Johannesburg, South Africa.

Corresponding author:

Mr N Bacci, MSc

Human Variation and Identification Research Unit, School of Anatomical Sciences, University of the Witwatersrand, 7 York Road, Parktown 2193, South Africa

Email: nicholas.bacci@wits.ac.za and nicholasbacci@gmail.com

Tel: +27-11-717-2078

ORCID numbers:

Mr Nicholas Bacci: 0000-0001-6858-7598
Dr Tobias MR Houlton: 0000-0003-1215-0869
Dr Nanette Briers: 0000-0002-8544-6049
Prof Maryna Steyn: 0000-0002-0215-9723

Declarations:

Funding The current study was conducted with support from the South African National Research Foundation (NRF) and the J.J.J. Smieszeck Fellowship from the School of Anatomical Sciences, University of the Witwatersrand (DAAD-NRF and J.J.J. Smieszeck Fellowship funds awarded to N. Bacci (Grant No.: 11858) and NRF funds awarded to N. Briers as part of the Improving Methodologies and Practices in Craniofacial Identification (Grant No.: CSUR160425163022; UID: 106031). Any opinions, findings and conclusions or recommendations expressed in this study are those of the authors and therefore the NRF and University of the Witwatersrand, Johannesburg, do not accept any liability in regard thereto.

Conflicts of interest/Competing interests: All authors affirm that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Ethics approval: The study was unconditionally approved by the Human Research Ethics Committee (Medical) of the University of the Witwatersrand, Johannesburg. Ethics clearance certificate number: M171026.

Consent to participate: All participants included in the study signed an informed consent form granting the authors the use of their data.

Consent for publication: No participant information or photographs are published as part of the study, as agreed upon in the aforementioned informed consent signed for participation.

Availability of data and material: The datasets analysed during the current study are not publicly available due the sensitive nature of the Wits Face Database. The database has the identifier http://dx.doi.org.10.18772/10539/20690. A sample of the database (consisting of one of the author's' facial images and recordings) is made available at https://hdl.handle.net/10539/29924. Restrictions apply to accessing the database or part thereof, which was used under license for the current study and, hence, it is not publicly available. Access is limited to formal request and conditional approval by the School of Anatomical Sciences Collections Committee, strictly for non-commercial research.

Code availability: Not applicable

Authors' contributions: Conceptualization: Nicholas Bacci, Maryna Steyn, Nanette Briers; Methodology: Nicholas Bacci, Maryna Steyn, Nanette Briers, Tobias Houlton; Formal analysis and investigation: Nicholas Bacci, Tobias Houlton; Writing - original draft preparation: Nicholas Bacci; Writing - review and editing: Nicholas Bacci, Maryna Steyn, Nanette Briers, Tobias Houlton; Funding acquisition: Nicholas Bacci, Maryna Steyn, Nanette Briers; Resources: Maryna Steyn, Nanette Briers; Supervision: Maryna Steyn, Nanette Briers

Acknowledgements

Thanks are due to all the participants who agreed to be photographed and recorded for the development of the Wits Face Database from which the face pool samples were composed. A special thanks is also due to Joshua Davimes for his instrumental contributions in developing the Wits Face Database. We are also especially grateful towards Tamara Lottering for her assistance in composing the face pools and Gideon LeRoux for his assistance in the capturing and extracting of the CCTV recordings from the University security systems. Thanks are also

due to the volunteers who aided in participant recruitment: Jesse Fredericks, Kiveshen Pillay, Rethabile Masiu, Sameerah Sallie, Daniel Munesamy, Laurette Joubert, Jordan Swiegers, Betty Mkabela, Johannes P. Meyer, Amy Spies, Natasha Loubser, Nicole Virgili, Dan-Joel Lukumbi, Tamara Lottering, Mathabatha Ntjie, Claudia Landsman, Raheema Dalika, Merete Goosen, Stephanie Souris, Rabelani Negota, Mahlatse Mahasha, Jessica Manavhela.

Abstract

Between the ever-increasing availability of surveillance evidence and expert-based forensic facial comparison being considered admissible in court, confirming its validity is paramount. Facial comparison is most commonly conducted using morphological analysis (MA), a largely untested feature-based approach. This study aimed at validating the current recommended practice of MA in both standardised and suboptimal surveillance samples. Face pools of 175 South African males were compiled with a series of facial photographs, using images from the Wits Face Database. The first 75 face pools consisted of wildtype (unstandardized) high-quality target photographs, while the remaining 100 face pools consisted of suboptimal CCTV target images. Target images were compared to high-quality standardised photographs. Face pools were analysed using the Facial Identification Scientific Working Group's guidelines and feature list. Confusion matrices were used to determine the performance of MA in each cohort. MA was found highly accurate (chance-corrected accuracy (CCA): 99.1%) and reliable (κ=0.921) in the photographic sample and less accurate (CCA: 82.6%) and reliable (κ = 0.743), in the CCTV sample. Higher false positive and false negative rates were noted for the CCTV sample, with the majority of errors resulting in false negative outcomes. The decreased performance in the CCTV sample was attributed to various factors including image quality, angle of recording, and lighting. Other studies testing facial comparison identified lower accuracies and reliability across various conditions. Better performance was found here and in other studies that included some form of facial feature list, reinforcing the importance of using a systematic facial feature list.

Keywords: forensic facial comparison, face mapping, facial identification, morphological analysis, CCTV, FISWG.

Introduction

Forensic facial identification makes use of photographic or video evidence of a target's face to confirm his/her identity and/or possible association with a criminal event. Historically, forensic facial comparison (FFC) has been employed by experts brought forth in defence of suspects to contest image-based evidence, in order to exclude a suspect based on inconsistent facial features [1]. In the last 20 years, expert testimony has been deemed admissible for conviction, provided that images are of good quality [1]. Several cases have also served in South African courts where FFC's have contributed towards successful convictions. This highlights the importance of the use of high-quality image and recording in the process of the analysis,

especially as training and experience may not particularly aid in correctly identifying unfamiliar faces from "very poor quality" CCTV recordings [2, p. 25].

Due to the increase in available video and photographic evidence, FFC, or face mapping, is now a commonplace practice in many judicial systems [3]. While efforts towards automated and semi-automated facial recognition systems are on the rise both in the private and government sectors [e.g. 4], the resulting methods and systems seem to be effective only in highly standardised settings [e.g. 5]. This standardised setting requirement primarily results from the reliance of these systems on facial anthropometric information to accurately compare faces [6, 7]. Despite recent attempts at standardisation [8], photo-anthropometry, the application of facial anthropometry to image comparison, has been demonstrated as unable to discern target individuals for both identification and exclusion [9, 10]. This is further complicated by the requirement of the application of photogrammetry to extract dimensions on many images in order to affect a comparison, something not possible if the images are not highly standardised [11] or of good enough quality [9]. As a result, the recommended method to perform individual facial comparisons for criminal or legal proceedings is morphological analysis [9, 12].

Morphological analysis is regularly used in the judicial context across the world despite being universally considered highly subjective in nature [11] and the limited testing of its accuracy or efficacy [12, 13]. In fact, no other experimental studies have tested the accuracy of facial comparison by morphological analysis [11]. Morphological analysis is based on a face analysis expert's opinion on the extent of facial similarity between unfamiliar faces [14]. In circumstances involving familiar faces, even without the use of morphological analysis, eyewitnesses and other previously familiarised untrained persons are considered adequately able to identify individuals even in poor quality images and CCTV footage [15, 16]. However, the public has been assessed as generally poor at identifying unfamiliar faces [16–18] and eyewitness evidence has been disputed when involving unfamiliar faces [19, 20]. Testing of the performance of broad concepts of FFC has been conducted, but primarily focused on the influence of experience of analysts to achieve the best results [2, 21] or to test the psychological interactions involved in facial identification [e.g. 18, 22, 23]. For FFC experts to undertake the process of morphological analysis correctly, it is recommended that a systematic comparison be done to compare a specific list of features in every analysis [12]. This feature-by-feature comparison of the face between a target image and a reference sample can be done following any of multiple feature lists available [12, 24, 25]. While recent studies have shown both feature

instructions and feature by feature comparisons to benefit the accuracy of facial comparisons [26, 27], specific feature lists have not been extensively tested. Although some limited testing of the efficacy of facial comparison using facial feature lists has been done [24, 28], the most extensive, yet largely untested [13], list of facial components and descriptors was published by the Facial Identification Scientific Working Group (FISWG) [29]. As such, a detailed test of feature-by-feature comparison based morphological analysis is required to confirm the accuracy and validate the use of this approach in forensic facial identification.

The aim of this study was therefore to use the prescribed FISWG list of morphological features, in order to assess their ability to accurately identify unknown individuals on both unstandardized photographs and CCTV images. For this purpose, standardized photographs from a large database of South African individuals were compared to unstandardized photographs and selected images from CCTV footage within face pools.

Materials and Methods

This project forms part of a larger study. All images were drawn from the Wits Face Database [30]. The sample included facial photographs and images extracted from CCTV recordings of a total of 175 consenting, known individual participants. All participants were young male South African adults of African descent. These photographs were captured at a height of 1600 mm, resulting in an effective angle of incidence of 0° in relation to the participant's face. The standardised photographs (ST) were captured using a Canon 1300D 18MP DSLR camera (18-55 mm DC Canon lens) at an objective to face distance of 1500 mm using a black velvet cloth backdrop and cloth to obscure the participants' clothing. The unstandardized "wildtype" photographs (WT), intended to simulate suboptimal photographs but with a similar quality camera, were captured using a Sony SLT A57 (18 – 250 mm Sony Zoom Lens) at a distance of 5000 mm. These photographs were captured at five views or facial poses: anterior, right 45-degree, right lateral, left 45-degree and left lateral view. The image resolution for the ST photographs and WT photographs was 3456 x 4320 pixels at 300 dpi and 3264 x 4080 pixels at 350 dpi, respectively. CCTV recordings were extracted from an Internet Protocol (IP) camera (HikVision, model: DS-2CD2142FWD-I, 4-megapixel) previously installed for surveillance at the University of the Witwatersrand. The objective to face distance of the CCTV camera was approximately 3030 mm and the angle of incidence was approximately 27°, calculated based on the mean height of the study's population (South African males) of 1708 mm [31]. Still images were obtained from the CCTV recordings of the matching individuals

at the same five views captured in the photographic sample. These images and photographs were combined into face pool trials based on broadly visually similar individuals to test the efficacy of the FISWG facial feature list for individual identification.

This study included a total of 175 face pool trials, each composed of a single target facial photograph grouped with either nine foil images and one correctly matching image (test trials) or 10 foil images (no-match trials). The foil images were assigned to face pools based on relative facial resemblance and apparent age. This was done by a research assistant, to ensure blinding the primary analyst. These face pools were grouped by the target type of facial images and arranged in corresponding views or facial poses. The first 75 face pools included a "wildtype" (WT) high-resolution target photograph, each compared to 10 facial images from a database composed of standardised (ST) optimal photographs with obscured clothing. The remaining 100 pools were composed of a captured frame from standard CCTV recordings and 10 ST facial images. Throughout all facial comparisons, each target image was compared to a single photograph from the 10 photographs of the grouped face pool at a time. The CCTV extracted images were cropped at the same aspect ratio (4:5) as the photographic images with the face centred in the same manner, resulting in the images having an approximate resolution of 368 x 460 pixels at the native 96 dpi of the CCTV camera. In this manner the face represented approximately $1/9^{th}$ of the entire image, resulting in the face being composed of approximately 41 x 52 pixels. A limited number of maximum 15 repeat uses per ST photograph was allowed while compiling face pools to prevent familiarisation of the principal analyst.

Each face pool included a relatively even distribution of views, or facial poses, of the target and corresponding faces (Table 2). Different facial poses were included in the study to allow for a varied review of the full face of the individuals compared. The even distribution of facial poses throughout the composition of the face pools allowed for all parts of the feature list to be used and explored throughout this analysis. Including a single view only, although making the power of the analysis stronger, would limit the facial features reviewed and would not be representative of the entire face. In turn, this permitted a comparison of the efficacy of the feature list across the varied poses.

It is important to ensure that the observations are reliable and repeatable. Among these face pools, repeat trials were thus included to allow for intra-observer error calculation. These repeat trails were distributed within the analysis sample so that the first original appearance of the same face pool and the order of the photographs within the face pool were distant enough

to dismiss familiarisation. The sample breakdown for each analysis cohort is detailed in Table 1.

In order to assess inter-observer agreement, a facial comparison expert performed a series of facial comparisons in a subset of five face pools from each cohort (total 10 face pools, including 100 total individual facial comparisons). This was then to be contrasted to the first set of analyses by the primary investigator for inter-observer agreement calculation.

Facial comparison was conducted using FISWG's guidelines on facial comparison. To replicate a realistic scenario, each of the potential matches from a face pool was first briefly examined by the analyst using holistic review according to FISWG guidelines [12]. Holistic review was conducted as a brief overview of the entire face, with a self-imposed and controlled time limit of five seconds to prevent detailed review of the facial features. A decision was then made on whether the individuals could be a match or not, or if comparison was not possible. Holistic comparisons were scored as a prospective possible match (1), not a match (2), or no analysis possible (3). Only the prospective matches were further compared using morphological analysis. Morphological analysis (MA) is performed similarly to using an analysis, comparison, evaluation and verification (ACE-V) approach [12]. ACE-V is the commonest methodology employed in fingerprint examinations and consists of a systematic staged process that has been likened to the scientific method [32]. When a forensic analyst applies an ACE-V methodology, they will firstly analyse the anatomy and detail of the data available, followed by a comparison between the sets of data available, and lastly evaluate the information obtained from the comparison to come to a conclusion [32]. The verification phase then includes the confirmation by an additional analyst who retests the original conclusion [32]. The verification phase was not conducted across all analyses in the present study due to time constraints. MA comparison was conducted using the FISWG facial component feature and descriptor list [29] summarised in Table 3.

Morphological analysis was carried out by placing both images next to one another in Microsoft PowerPoint (365, v. 16) and adjusted to resemble the rotation and size of one another while maintaining the original aspect ratio. Morphological features of the face were then compared and contrasted for similarities or dissimilarities (Figure 1). These were highlighted in numerical order based on the FISWG facial component list and used to make a final decision for each possible match. Final decisions for each facial comparison were done using an adapted version of the levels of agreement system used by the South African Police Service [14].

The classifications and definitions of the adapted levels of agreement and summarising statements employed by South African Police Services (adapted from Steyn *et al.* [14]) are described below:

1. Conclusion: positive

   Considering the correspondence of general class characteristics and the presence of individual identifying unique characteristics that were found during the analysis between the person in the Target Images and in the Control Photographs, it is found that the person in the Target Images and in the Control Images probably depicts the same person.
2. Conclusion: inconclusive (but with some correspondence in a few visible morphological features)

   No conclusive findings regarding identification/individualization are possible because during the assessment the quality of the Target Images was found to be too poor to distinguish sufficient identifiable and unique characteristics. However, a few morphological similarities were visible between the person in the Target Images and in the Control Photographs. Because of insufficient evidence the person in the Target Images and the database photographs could not be identified as the same person, but neither could they be eliminated as not being the same person.
3. Conclusion: inconclusive (but with some correspondence in holistic appearance)

   According to a holistic comparison, there are corresponding general characteristics visible between the person in the Target Images and in the Control Photographs. However, due to the poor quality, light levels, angle and distance of the Target cameras, the Target footage is too poor to mark out sufficient specific morphological facial landmarks. Therefore, no detailed facial comparison could be performed between the Target Images and Control Images. No conclusive findings regarding identification are therefore possible because the Target Images were found to be too poor.
4. Conclusion: negative

   Considering the lack of correspondence of general class characteristics and individual identifying unique characteristics that were found during the analysis between the person in the Target Images and in the Control Photographs, it was found that the person in the Target Images and in the Control Images is not the same person.
5. Conclusion: no analysis possible

The images from the Target footage were of poor quality. Due to the low quality of the images and poor angle of camera or distance of camera, no sufficient morphological facial landmarks could be identified and therefore no facial comparison could be done.

In order to facilitate statistical analysis, these possible conclusions were adjusted for hit rate calculation as either being a correct match (scores 1, 2), not being a match (scores 3, 4), or an inconclusive analysis (score 5). These conclusions, which included the trials excluded as "not matching" in the holistic review phase of the analysis, were then compared to the true matchup of the images to calculate the hit rate of the principal investigator. The results of each comparison were as such categorised, based on the true matchup, as either a true positive (TP), a false positive (FP), a false negative (FN), a true negative (TN), or an inconclusive analysis (IA).

All statistical analyses were conducted with the R Studio statistics programme (version 3.6.3) [33]. Accuracy of facial identification for each group type of trials was obtained through hit rate calculation and chance mitigation by conducting a weighted Cohen's kappa. The target true accuracy for the method to be considered successful was a kappa value of 0.80 or above [34]. Intra- and inter-observer agreement were calculated using weighted Cohen's kappa with regard to the match up decisions.

From the hit rate of the comparisons, a full confusion matrix including multiple measures of predictive power was calculated from a contingency table (Table 4). The confusion matrix function in R calculates from the contingency table a raw overall accuracy by dividing the total correct scores by the total sample, along with a 95% confidence interval. This overall accuracy is then compared to a No Information Rate (NIR), or the largest proportion of the observed classes [35], from which a p-value is calculated to confirm whether the accuracy is higher than the NIR as an indicator of a good predictive model [36]. A McNemar's Test is also calculated as part of the function to determine whether the scored matches have a similar or different proportion of errors in the testing cohort [37], where a p-value below $\alpha = 0.05$ indicates a difference in proportion of errors. This comparison effectively indicates whether the analysis method errs relatively equally in false positives and false negatives.

The series of formulae used to calculate the quality of predictive power for morphological analysis within each subsample cohort were drawn from the Caret R package [35, 36], are defined below and presented in full as supplementary information (SI). Sensitivity, or true

positive rate or recall, is defined as the measure of how effectively a test can perform correct matches (SI: Equation 1). Specificity, or true negative rate, is a measure of how effective a test can correctly discard a match (SI: Equation 2). Prevalence is the total number of truly correct matches in the entire cohort sample (SI: Equation 3). The positive predictive value (PPV) indicates the probability of a trial scored as a match being truly a correct match (SI: Equation 4) and negative predictive value (NPV) is the probability of a trial scored as an exclusion being truly a correct exclusion (SI: Equation 5). The detection rate is the proportion of the cohort sample correctly matched (SI: Equation 6) and the detection prevalence is the proportion of scores matched by the analyst throughout the analysis cohort (SI: Equation 7). Ideally in this sample, the detection rate would be as close as possible to 9 or 10% due to how the sample was organised. An adjusted accuracy calculation that considers the unbalanced distribution of the sample is the balanced accuracy calculated using equation 8 (SI) making use of sensitivity and specificity. The ratio of correctly assigned matches to total positive matches made by the analyst is termed precision and calculated using equation 9 (SI). Lastly, the F-Score is a weighted (harmonic) mean of Precision and Sensitivity calculated using equation 10 (SI) and considers both false positive and false negative designations. This is calculated by adjusting the weight of Precision and Sensitivity in the equation; in this study equal weighting was given to both. As such in equation 10, β is assigned a value of 1 and the score is termed an F_1-Score.

False positive and false negative rates for all cohorts of analysis were also calculated separately to highlight the type of errors associated with each cohort. All the above-described analyses were conducted at two different thresholds to compare the performance of morphological analysis within a lenient and a strict iteration. The lenient iteration was described previously including comparisons scored as 1 and 2 as matches and 3 and 4 as not matching, however, the strict iteration includes strictly scores of 1 as positive matches while consolidating scores of 2 to 4 and treating them as not a match. This was intended to allow for only the highest level of confidence-based matches to be considered, in order to attempt reducing false positive outcomes. Chi-squared tests were used to compare the occurrence of true positive matches, false positive matches, false negative matches, true negative matches, or inconclusive analyses within a group type of trials across the different facial views (i.e. anterior, left lateral, etc.).

Results

Comparison of intra- and inter-observer agreement in face pools analysed by morphological analysis (MA) indicated a good to high level of agreement. Intra-observer agreement was

higher for MA ($\kappa = 0.934$; $p < 0.001$) in the photograph-to-photograph comparisons over the CCTV to photograph comparison ($\kappa = 0.807$; $p < 0.001$) (Table 5). Inter-observer agreement in the photograph-to-photograph trials was almost perfect ($\kappa = 0.921$; $p = 0.001$) (Table 5), but CCTV to photograph trials showed a slight decrease in agreement but still with a substantial level of agreement ($\kappa = 0.743$; $p < 0.001$) (Table 5).

A summary of the outcomes of each face pool is outlined in Figures 2 and 3 for both the photograph-to-photograph (Fig. 2) and CCTV to photograph comparisons, respectively (Fig. 3). No facial comparisons from the photograph-to-photograph trial cohort were classified as inconclusive and no trials were excluded from further analysis from this cohort (Fig. 2). A total of 26 potential matches could not be completed and were classified as inconclusive analyses in the CCTV analysis cohort and excluded from analysis (Fig. 3). This was due to a combination of poor quality and bad lighting in some of the CCTV recordings.

Confusion matrices confirmed a better overall performance of MA within the photograph-to-photograph sample. This better performance is highlighted in the contingency tables for each cohort of analysis (Tables 6 and 7). The raw overall accuracy across analysis cohorts was markedly higher for photograph-to-photograph MA (0.997, CI = 0.989 – 0.999) over CCTV, although the CCTV cohort MA's were also quite accurate (0.942, CI = 0.925 – 0.956) (Table 8). Raw accuracy is of limited use in analysing the performance of MA as it does not take sample composition and chance into account as it is a direct hit rate calculation. All analysis groups had a greater accuracy than the No Information Rate (NIR) (Table 8), which is indicative of a good predictive model of morphological analysis overall.

Chance corrected accuracies (CCA) of MA indicated a decrease in predictive strength from photograph-to-photograph comparisons ($\kappa = 0.982$; $p < 0.001$) to CCTV comparisons ($\kappa = 0.710$; $p < 0.001$) (Table 8). The accuracy for the CCTV cohort remained below the considered "almost perfect" score of 0.80 [34]. Within the photograph-to-photograph cohort there was no statistically significant proportion of errors (Table 8), which is indicative of the markedly low false positive and false negative outcomes (Tables 6 and 7). An increased false positive rate (Table 9) is evidenced by the higher detection prevalence against the detection rate of the CCTV cohort of analyses, resulting in a markedly decreased precision (Table 9). An F_1-Score lower than 0.8 is also seen as a result of the higher false positive rate across the CCTV analysis (Table 9).

A second, strict, iteration of statistical analyses was conducted by adjusting the threshold of correct matches to only include positive conclusions (score of 1). This stricter interpretation was intended to reduce the risk of false positive outcomes. In this iteration, chance corrected accuracies were higher for both photograph-to-photograph ($\kappa = 0.991$; $p < 0.001$) and CCTV ($\kappa = 0.826$; $p < 0.001$) cohorts than the previous, lenient iteration of analysis. The increase in accuracy results from the strict iteration correctly assigning some of the false positive matches present in the lenient iteration to true negative matches. Confusion matrices also demonstrated a marked decrease in false positives (Tables 10 and 11) and a slight increase in false negatives in the CCTV cohort (Table 11). Raw accuracy of prediction also seemed to increase based on the confusion matrix (Table 12). The lack of false positive outcomes and extremely low overall false outcomes in the photograph-to-photograph cohort (Table 10) resulted in the McNemar's test indicating no statistically significant proportion of errors (Table 12). The shift in false positives and negatives was evidenced by the decrease in sensitivity in the CCTV cohort and particularly the increase in specificity across both cohorts in the strict analysis iteration (Table 13). This decrease in sensitivity in the CCTV cohort results in a slightly diminished balanced accuracy. However, it also resulted in a markedly higher precision and F_1-Score (Table 13), indicating an overall better performance of the strict scoring interpretation and highlighting this as the ideal approach for forensic applications. False positive rates (FPR) and false negative rates (FNR) calculated across the analysis cohorts indicate a higher FPR and a higher FNR in the CCTV cohort over the photographic only cohort (Table 14), demonstrating the overarching benefit of the use of higher-quality images in facial comparison. When considering the strict iteration, while a notable decrease in FPR was noted in the CCTV sample, a parallel doubling of the FNR was also identified (Table 14). Across all trials, balanced accuracy, precision, and F1-Scores were higher in the photograph-to-photograph sample.

Lastly, Chi-squared tests comparing the performance of MA across different views demonstrated no statistically significant difference in performance for any particular facial view. This result indicated that no real relationship can be determined with obtaining a higher number of correct matches in a specific facial view. This suggests that provided enough features are visible in suboptimal CCTV images, each view can provide enough information to compare faces accurately.

Discussion

The common use of facial photographic and video surveillance evidence in the legal system [38, 39] has led to an increased need for forensic facial comparison (FFC). Testing the validity of FFC is thus crucial for admissibility of this evidence and accurate legal processes. FFC employed on an unfamiliar face sample is a grossly understudied field in craniofacial identification, with most studies focusing on the benefit of expertise in FFC [e.g. 2, 21]. Confirming facial comparison experts' work as valuable and admissible is vital for the consideration of their reports. However, it is important to prevent this expertise-based evidence being integrated in court without confirmation on the accuracy of this method [39]. As such, the validity of the most accepted and utilised approach to FFC, viz. morphological analysis, needs to be tested and confirmed.

The present study is the first to extensively test the accuracy of morphological analysis (MA) in an actualistic scenario, simulating conditions encountered by experts in the forensic setting. The efficacy of MA in this actualistic sample resulted in remarkably high accuracies, ranging from 71% to 99%, depending on sample type and strictness of analysis. Moreover, a particularly high level of intra- and inter-observer agreement was identified across cohorts as well, ranging from substantial ($\kappa = 0.743$) to almost perfect ($\kappa = 0.934$). A noticeably better performance was observed in the photographic sample, both in terms of accuracy (raw accuracy (RA): 99.7%; balanced accuracy (BA): 99.2% chance corrected accuracy (CCA): 96.8%) as well as intra- and inter-observer agreement ($\kappa = 0.934$ and $\kappa = 0.921$, respectively). This outcome was expected, considering the markedly higher quality and better angle of view correspondence of standardised and wildtype photographs. The correspondence and higher quality of the images in turn allowed an in-depth facial component review and therefore, a more informed decision to be made. When MA failed to identify matches correctly, it predominantly resulted in false negatives in both photographic (1.61%) and CCTV (11.63%) samples. False positive outcomes were, in fact, notably lower than false negatives across both samples (photographic: 0.16% and CCTV: 5.19%).

False positives and false negatives are considered as sensitivity and specificity when comparing test method performance. In a practical setting, the predictive cut-off of any test is scaled to obtain the optimally acceptable ratio of sensitivity and specificity. This was not entirely possible in this study as the outcomes of the facial comparisons are assigned based on a limited ordinal four-scale scoring system, that could only be reasonably adjusted at two levels.

However, by conducting a lenient and strict iteration of the data analysis, it was possible to control, within limits, the performance of morphological analysis to identify the optimal iteration to be used in the forensic setting. In the strict iteration, a drastic reduction of false positives was evident for both the photographic sample (from 0.16% to 0%) and CCTV analysis cohorts (from 5.19% to 0.60%). This low rate of false positives is indicative of the importance of a scale for levels of agreement with distinct criteria such as the one employed in the present study [14], which can allow for assigning final conclusions in a stricter manner than the previously suggested the Bromby [40] scale of support. The downside of this stricter approach is the near doubling of the false negative rate found in the suboptimal images of the CCTV sample (from 11.63% to 23.26%). This higher rate of false negative outcomes identified could potentially be partially attributed to the effect of distinctiveness. Faces that are composed of more distinct or unique features are intrinsically easier to match [42, 43] or discard, while faces that are more typical with more "usual" feature types are more likely to be recognised [42], which would result in them being assigned as potential matches. While the scale adjustment used in the current study resulted in higher false rejections than other studies (10.5% and 8%) [2, 21], it should be considered in the medicolegal context where the consequences and implications of accepting a methodology with a higher false positive rate can be dire. Wrongful convictions of factually innocent individuals can result from high false positive rate-based methods, and it is generally accepted that a lower false positive rate should be prioritised in acceptable and admissible forensic methodologies [41].

Beyond the marked reduction in false positive rates at the expense of higher false negative rates, the strict iteration also yielded both higher raw accuracies (99.9 and 97.3%) and chance corrected accuracies (99.1 and 82.6%) in both sample cohorts. Balanced accuracy decreased slightly in the CCTV sample (88.1%) as a result of the decreased sensitivity (0.767) of the strict iteration resulting from the higher rate of false negatives. Even the positive predictive values were higher for the strict iteration across both cohorts (1.000 and 0.930) with relatively constant negative predicting values (0.998 and 0.977). Most importantly, the highest precision (1.00 and 0.930) and F_1-scores (0.992 and 0.841) were achieved in the strict iteration as well, indicating the overall better performance of a strict model. Due to its decreased chance corrected accuracy, and decreased precision and F-scores, the lenient model should only be applied in a forensic context when other supporting evidence is available. Similarity in certain morphological traits (score of 2) should not, on its own, be considered by the prosecution to

promote a strong likelihood of a defendant being the target in question, as those scores could sometimes be false positives.

Irrespective of iteration, the CCTV cohort resulted in an uneven proportion of errors, with a tendency towards false positives in the lenient iteration and false negatives in the strict iteration. In the photographic sample, no false positives were found with the strict analysis and the false negative rate persisted at the original low levels (1.61%). In a forensic setting, this is a preferable outcome, as it is important to err on the side of the accused so that potentially innocent individuals are not linked to a crime.

In summary, across all metrics used to determine the efficacy of MA, it was clear that MA performed better in the optimal and standardised setting of the photograph-to-photograph comparisons. This is demonstrated by the extremely high precision, sensitivity, specificity, positive and negative predictive values identified (ranging between 0.984 and 0.998). These results are indicative of the fact that, despite its subjective nature, under ideal conditions, MA is able to identify correct matches and reject false matches with a very high probability of being correct. However, when taking into consideration suboptimal images, which is often what forensic image analysts are presented with [44], the performance of MA declined. Notably, in the suboptimal CCTV cohort, MA was found to be less effective in correct matching (RA: 94.2%; BA: 91.4% CCA: 73.6%) and less reliable (intra-observer $\kappa = 0.807$ and inter-observer $\kappa = 0.743$ agreement), while almost identically capable of discarding incorrect matches. The notable lower positive predictive value (PPV) observed highlighted the decreased probability of a match being correct. As the decrease in negative predictive value in the CCTV sample is minor by comparison, the effect on the PPV could have been magnified by the study design and the organisation of the sample analysed. The sample was arranged so that the majority of the comparisons are true negatives, as shown by the prevalence of 9.4%, representing the percentage of the sample that can be correctly attributed as true positive. As a result, a change to how positive (true or false) cases are assigned, would be more evident. Interestingly, the increase in detection prevalence (0.130), coupled with the decrease in detection rate (0.083), may be indicative of an increased likelihood to falsely assign matches, while overall assigning more matches, when the image conditions are suboptimal.

Incorrect matching could have resulted from a multitude of factors including image quality, photographic/recording conditions, lighting and position of the cameras. The photographic images used in the current study, with their much higher resolution (3456 x 4320 pixels, 300

and 350 dpi), were as such considered conducive to more accurate analyses than the drastically lower quality (358 x 460 pixels, 96 dpi) CCTV footage. Notably, the quality of footage used in this study was considered as high-resolution (>2 Megapixel) and of a pixel density well over the recognised standard of approximately 25 dpi required for adequate facial analysis [45]. Lee and colleagues [2] found accuracies of 67% and 67.3% for untrained and trained assessors, respectively, when using poor quality footage for facial comparisons. This lower-end hit rate across both experts and the public further emphasizes the necessity of high-quality images, despite training. Other investigations into higher-quality CCTV based facial on comparison reported 46% and 83% correct identifications by the public and experts, respectively [21]. The accuracy reported by Wilkinson and Evans [21] for experts is in a similar range to the accuracies reported for the CCTV sample in the present study, depending on the calculation for accuracy conducted. In addition, Wilkinson and Evans [21] identified that facial analysis experts had a higher false acceptance (17%) than false rejection rate (8%). Lee *et al.*, [2] found that false acceptance in poor quality CCTV footage was twice (22.4%) as likely as a false rejection (10.5%). Comparatively, the present study identified drastically lower false acceptance (false positive) rates of 0.15%, for the photographic sample, and 5.22%, for the CCTV cohort, and varying false rejection (false negative) rates of 1.43%, for the photographic, and 11.96%, for the CCTV cohorts. These outcomes demonstrate the importance of good quality of images, which has also been previously reiterated by other researchers, when expert testimony was deemed admissible for conviction, provided that images are of good quality [1].

Beyond the quality of the images analysed, the position of the CCTV camera may have also influenced the ability of the analyst to compare facial features accurately, as the part of the face that is visible and the angle of the face are altered. Standardised photographs were captured at a perpendicular angle, while the CCTV was recording the participants from a greater height and with an angle of incidence around 27°, as is the case in real-life situations. The maximum angle of incidence recommended for facial detail capture in CCTV is 15° [46]. This angle is recommended particularly as some individuals tend to have a natural inferior head tilt of 15 to 20° while walking [46]. In the current study, however, participants stood still facing straight ahead in the five views captured, which mitigated the natural head tilt. Despite this, the 27° angle reduced visibility and may have altered proportions of facial features (Fig. 1) enough to contribute to the increased difficulty in facial comparison. Limited to non-existent visibility of the eyes was noted in most CCTV images, irrespective of face view, as generally identified in situations with an angle of incidence greater than 15° [46].

In addition to the angle and overall raw image quality as shortcomings of CCTV in facial comparison, lighting may often be suboptimal to clearly visualise all facial features. Standardised lightning in the photographic sample allowed for clear comparisons, while the natural sunlight in the CCTV sample illuminated the participants in an uneven manner (Fig. 1). Uneven illumination commonly results in large contrast shifts in the subjects recorded [47]. This contrast shift may result in overexposure of subjects directly under the light source with the camera not being able to attenuate the large contrast variation, causing loss of detail [47]. In the present study's CCTV sample, these lighting effects were clearly noticeable (Fig. 1) and affected the ability of the analyst to compare details effectively.

Despite these differences in the conditions of image acquisition, morphological analysis performed well, further suggesting that variation of conditions, within reason, does not undermine facial comparisons [11]. In addition, irrespective of the difficulties created by the mismatching image acquisition conditions, facial poses or views did not seem to affect the performance of morphological analysis across both photographs and CCTV as no differences were identified throughout the sample provided, since all faces were arranged in matching poses (i.e. left lateral standardised image versus left lateral CCTV extracted image). This seems to suggest that despite the large amount of variation in the facial features exposed and visible, enough features were observable and comparable between images of matching poses. It was also noted qualitatively during analysis that facial comparisons relied more on certain features, dependent on the view used. For example, the ear and nose seemed to be unique and distinct enough in lateral images to provide the most distinguishing comparable features, while the mouth, nose and eyes (when visible) were preferred during anterior view comparisons. The effect of pose variation has been extensively investigated in computer-based facial recognition as one of its greatest challenges to overcome, with some promising prospects for future research [48]. This is evident when even the famed eigenface approach to facial recognition modelling would require more specific inclusion of varying viewpoints [49], a capacity thought intrinsic to human-based face recognition [50]. Despite this awareness, no other facial comparison studies have reported on the effects of various poses to facial identification, specifically in relation to a feature-based approach such as morphological analysis. As such, further testing in a larger sample with consideration of both matching and mismatching views could aid in confirming the results of this study.

Generally, reliability has been reported to be inconsistent for facial comparison and morphological facial feature lists particularly. A noteworthy facial feature list was established

by the Facial Identification Scientific Working Group (FISWG), one of the working groups of the organisation of Scientific Committees for Forensic Sciences, which operates in conjunction with the United States National Institute for Science and Technology. FISWG comprises of experts who aim to guide practice and techniques in the field. Using the FISWG feature list, the present study found a high agreement rate on a purely morphological feature analysis method (intra-observer: 93.4% and 80.7% for photographs and CCTV respectively; inter-observer: 92.1% and 74.3% for photographs and CCTV respectively). In comparison, an older study demonstrated low error margins with limited morphological features and the inclusion of some morphometric data (0.07%), even though the nature of this error margin is not specified [28]. Others indicated much higher error margins despite detailed morphological feature lists (intra-observer error: 19% and inter-observer error: 39%) [24].

Considering the high reliability and the decrease of false positives found in the present study, a strong argument can be made for the use of a facial feature list when conducting MA. This recommendation is in agreement with recent studies that highlight the importance of experienced observers using a systematic approach, relying on a consistent feature list across all analyses, to achieve a high level of reliability in facial comparison [24, 51]. Establishing an accurate and reliable approach to facial comparison is paramount to decrease the risk of wrongful convictions is a major concern in the legal system.

Despite the importance of the systematic approach advised by FISWG and others [12, 24, 51], only a handful of studies investigating facial comparison developed, tested or reviewed specific feature lists [24, 28, 52], with the most detailed and recent feature list published [29] remaining untested until the present study. A great benefit of the FISWG feature list, other than its extensive nature, is the inclusion of descriptors to guide the forensic analyst in their efforts, an aspect lacking from other atlases where definitions or descriptors tend to be simplistic or assumed [24, 28, 52, 53]. Some of these studies primarily reported the prevalence of these specific feature morphologies identified within their sample and limited their discussion to which features may be most useful for discrimination in facial comparison without testing them in that specific ambit [e.g. 47, 48]. A recent study found an improved performance of face matching tasks when making use of facial feature instructions [27]. Although the observers were not expert facial analysts and the feature instructions were not as extensive as the FISWG feature list, better performance was seen when comparing faces of a similar population group to the observer [27]. Megreya and Bindemann [27] also stressed the importance of selection of specific features that may increase accuracy over others, citing the eyebrows as a particularly

beneficial feature to observe – more so than the eyes and ears. Similar findings were reported by an investigation into direct feature by feature comparisons [26]. According to Towler and colleagues [26], facial comparisons using features yielded higher accuracies with trained forensic analysts, to almost perfect identity discrimination when multiple analysts' aggregate scores were considered.

The stricter approach employed in the present study was intended to be in alignment with the legal principle of presumption of innocence that a vast number of United Nations recognised territories abide by in criminal proceedings (Article 11, United Nations Universal Declaration of Human Rights and Article 14 (2) International Covenant on Civil Rights). South African law also constitutionally abides by this principle (section 35(3)(h) of the 1996 the Constitution of the Republic of South Africa). Maintaining the presumption of innocence is a rising concern as many industries and governments are considering deployment of semi-automated facial recognition systems, with knowingly high rates of false positives [54–56] and racial biases [57, 58]. To mitigate the large number of false positive errors these systems are prone to, it will be crucial to rely on a low false positive rate human observer-based method during match screening as a contingency method. Preventing false positive conclusions is a matter of great ethical relevance, as a false match can result in innocent individuals being harassed or convicted [58]. As such, the strict iteration is recommended when surveillance data is the only evidence available, in order to prevent wrongful convictions. However, in circumstances where additional supporting evidence is available, the lenient iteration is still considered of great value in contributing towards building a strong case.

The lack of information on morphological analysis validation throughout the literature is understandable when taking into consideration the complexity of designing a truly representative study. A major challenge encountered in this and similar studies relates to the accessibility and availability of matching facial images of appropriate quality to permit facial comparison testing. While many databases exist, most are inconsistent in their image capture procedures and quality, while also having limited iterations or numbers of participants included [e.g. 54–56]. The present study relied upon the purposefully created Wits Face Database [30] with multiple matching known individuals, that lends itself well to testing facial comparison due to its standardisation. In addition, the requirement to arrange a series of blind face pools including multiple participants is a laborious task requiring independent external involvement. Creating blind face pools is required to prevent biasing the analysts, and in order to prevent familiarisation of the faces by controlling the number of times an analyst is exposed to them.

Over familiarisation needs to be controlled for, as familiar faces are intrinsically processed differently by humans [62–64] and would not be a truly representative test of forensic facial comparison. Familiarisation is not generally discussed in studies investigating facial comparison, as time limits or limited exposure to the face samples are not generally considered [e.g. 2, 28, 58]. However, in most studies of this nature, multiple analysts review the same photographs, limiting the concern of familiarisation [e.g. 2, 21].

Morphological analysis is also a time-consuming process, with the average analysis usually requiring over two hours [65], making the process of controlling exposure more complex. This is particularly troublesome when considering that performing facial comparison over prolonged periods of time, being an extensive and repetitive task, can severely hamper the accuracy of the analysts [66]. These are all factors that need to be considered not only when designing studies that investigate the accuracy of facial identification but also to assure a firm scientific approach to studies of craniofacial identification [11].

Conclusion

This is the first study to test the FISWG feature list for FFC in a large sample, within an actualistic context, specifically focusing on the accuracy that can be obtained by using this list of features. Excellent accuracies were found through both optimal and suboptimal testing conditions, with a high level of reliability. Low false positive rates were identified, suggesting that use of morphological analysis is accurate, reliable and has a low risk of false accusation when applied by experienced analysts in optimal conditions. A noticeable increase in false negatives was identified when only the highest level of confidence matches were considered as positive. However, this is likely an acceptable sacrifice, as preventing wrongful convictions by erring on the side of caution is widely considered a more acceptable practice in the judicial context. Overall, photograph-to-photograph comparisons performed better than CCTV to photograph comparisons, suggesting that strict criteria should be used when confirming a positive match in CCTV analyses.

With the predominantly good results, outperforming those of most other studies that did not use facial feature lists [2, 21], a strong argument can be made to justify the use of a feature list-based morphological analysis in the judicial context, in order to reduce subjectivity. As such, it is recommended to rely primarily on morphological analysis, based on an extensive feature list, when employing facial comparison in the medicolegal context. This recommendation is given with the provision that it should not be considered as the only evidence. It is imperative

to use further supporting evidence that can be associated with a facial analysis for conviction. This is particularly important in circumstances where analyses are considered inconclusive, despite some level of similarity (morphological or holistic).

To prevent wrongful convictions, strict iterations of the analysis should be considered as the preferred practice, to mitigate the risk of false positives. The more lenient iteration should only be used when other supporting evidence is available. Even though a higher false negative rate was found using the stricter approach, one can argue that this approach remains of benefit when used for exclusion in lower quality images to possibly eliminate potential suspects. A strict model application also aligns better with the necessary forensic standards for admissibility of evidence and fundamental human rights.

References

1. Davis JP, Valentine T, Wilkinson C (2012) Facial image comparison. In: Wilkinson C, Rynn C (eds) Craniofacial Identification, 1st ed. Cambridge University Press, New York, pp 136–153

2. Lee W-L, Wilkinson C, Memon A, Houston K (2009) Matching unfamiliar faces from poor quality closed-circuit television (CCTV) footage: An evaluation of the effect of training on facial identification ability. AXIS Online J Cent Anat Hum Identif 1:19–28

3. Jain AK, Klare B, Park U (2012) Face matching and retrieval in forensics applications. IEEE Multimed 19:20–27. https://doi.org/10.1109/MMUL.2012.4

4. Phillips PJ (2011) Improving face recognition technology. Computer (Long Beach Calif) 44:84–86. https://doi.org/10.1109/MC.2011.87

5. Urbanová P (2016) Performance of distance-based matching algorithms in 3D facial identification. Egypt J Forensic Sci 6:135–151. https://doi.org/10.1016/j.ejfs.2016.04.004

6. Davis JP, Valentine T, Davis RE (2010) Computer assisted photo-anthropometric analyses of full-face and profile facial images. Forensic Sci Int 200:165–176. https://doi.org/10.1016/j.forsciint.2010.04.012

7. Heyer R, Semmler C (2013) Forensic confirmation bias: The case of facial image comparison. J Appl Res Mem Cogn 2:68–70. https://doi.org/10.1016/j.jarmac.2013.01.008

8. Caple J, Stephan CN (2016) A standardized nomenclature for craniofacial and facial anthropometry. Int J Legal Med 130:863–879. https://doi.org/10.1007/s00414-015-1292-1

9. Kleinberg KF, Vanezis P, Burton AM (2007) Failure of anthropometry as a facial identification technique using high-quality photographs. J Forensic Sci 52:779–783. https://doi.org/10.1111/j.1556-4029.2007.00458.x

10. Moreton R, Morley J (2011) Investigation into the use of photoanthropometry in facial image comparison. Forensic Sci Int 212:231–237. https://doi.org/10.1016/j.forsciint.2011.06.023

11. Stephan CN, Caple JM, Guyomarc'h P, Claes P (2019) An overview of the latest developments in facial imaging. Forensic Sci Res 4:10–28. https://doi.org/10.1080/20961790.2018.1519892

12. Facial Identification Scientific Working Group (2019) Facial Comparison Overview and Methodology Guidelines. Available at: https://fiswg.org/fiswg_facial_comparison_overview_and_methodology_guidelines_V1.0_20191025.pdf

13. Houlton TMR, Steyn M (2018) Finding Makhubu: A morphological forensic facial comparison. Forensic Sci Int 285:13–20. https://doi.org/10.1016/J.FORSCIINT.2018.01.022

14. Steyn M, Pretorius M, Briers N, et al (2018) Forensic facial comparison in South Africa: State of the science. Forensic Sci Int 287:190–194. https://doi.org/10.1016/J.FORSCIINT.2018.04.006

15. Burton AM, Wilson S, Cowan M, Bruce V (1999) Research Article Evidence From Security Surveillance. Psychol Sci 10:243–248. https://doi.org/10.1111/1467-9280.00144

16. Bruce V, Henderson Z, Newman C, Burton AM (2001) Matching identities of familiar and unfamiliar faces caught on CCTV images. J Exp Psychol Appl 7:207–218. https://doi.org/10.1037//1076-898X.7.3.207

17. Burton AM, Wilson S, Cowan M, Bruce V (1999) Face recognition in poor-quality video: evidence from security surveillance. Psychol Sci 10:243–248.

https://doi.org/10.1111/1467-9280.00144

18. Henderson Z, Bruce V, Burton AM (2001) Matching the faces of robbers captured on video. Appl Cogn Psychol 15:445–464. https://doi.org/10.1002/acp.718

19. Megreya AM, Burton AM (2008) Matching faces to photographs: poor performance in eyewitness memory (without the memory). J Exp Psychol Appl 14:364–372. https://doi.org/10.1037/a0013464

20. Megreya AM, Memon A, Havard C (2012) The Headscarf Effect: Direct Evidence from the Eyewitness Identification Paradigm. Appl Cogn Psychol 26:308–315. https://doi.org/10.1002/acp.1826

21. Wilkinson C, Evans R (2009) Are facial image analysis experts any better than the general public at identifying individuals from CCTV images? Sci Justice 49:191–196. https://doi.org/10.1016/j.scijus.2008.10.011

22. Burton M, White D, McNeill A (2010) The Glasgow face matching test. Behav Res Methods 42:286–291. https://doi.org/10.3758/BRM.42.1.286

23. Megreya AM, Bindemann M, Havard C (2011) Sex differences in unfamiliar face identification: Evidence from matching tasks. Acta Psychol (Amst) 137:83–89. https://doi.org/10.1016/j.actpsy.2011.03.003

24. Ritz-Timme S, Gabriel P, Obertovà Z, et al (2011) A new atlas for the evaluation of facial features: advantages, limits, and applicability. Int J Legal Med 125:301–306. https://doi.org/10.1007/s00414-010-0446-4

25. Ritz-Timme S, Gabriel P, Tutkuviene J, et al (2011) Metric and morphological assessment of facial features: A study on three European populations. Forensic Sci Int 207:239.e1-239.e8. https://doi.org/10.1016/j.forsciint.2011.01.035

26. Towler A, White D, Kemp RI (2017) Evaluating the feature comparison strategy for forensic face identification. J Exp Psychol Appl 23:47–58. https://doi.org/10.1037/xap0000108

27. Megreya AM, Bindemann M (2018) Feature instructions improve face-matching accuracy. PLoS One 13:1–16. https://doi.org/10.1371/journal.pone.0193455

28. Porter G, Doran G (2000) An anatomical and photographic technique for forensic

facial identification. Forensic Sci Int 114:97–105. https://doi.org/10.1016/S0379-0738(00)00290-5

29. Facial Identification Scientific Working Group (2018) Facial Image Comparison Feature List for Morphological Analysis. Available at: https://fiswg.org/FISWG_Morph_Analysis_Feature_List_v2.0_20180911.pdf

30. Bacci N, Davimes J, Steyn M, Briers N (2020) Wits Face Database. In: Wits Institutional Repos. Environ. Dsp. https://hdl.handle.net/10539/29924

31. Steyn M, Smith JR (2007) Interpretation of ante-mortem stature estimates in South Africans. Forensic Sci Int 171:97–102. https://doi.org/10.1016/j.forsciint.2006.10.006

32. Speckeis C (2011) Can ACE-V be validated? J Forensic Identif 61:201–209. Available at: https://www.ncjrs.gov/pdffiles1/nij/225329.pdf

33. R Core Team (2020) R: a language and environment for statistical computing. Available at: https://www.r-project.org/

34. Watson PF, Petrie A (2010) Method agreement analysis: A review of correct methodology. Theriogenology 73:1167–1179. https://doi.org/10.1016/j.theriogenology.2010.01.003

35. Kuhn M (2020) Caret: classification and regression training R package. Available at: https://cran.r-project.org/package=caret

36. Kuhn M (2008) Building predictive models in R using the caret package. J Stat Softw 28:1–26. https://doi.org/10.18637/jss.v028.i05

37. Dietterich TG (1998) Approximate statistical tests for comparing supervised classification learning algorithms. Neural Comput 10:1895–1923. https://doi.org/10.1007/978-3-319-50926-6_6

38. Norris C, McCahill M, Wood D (2004) Editorial. The growth of CCTV: A global perspective on the international diffusion of video surveillance in publicly accessible space. Surveill Soc 2:110–135 https://doi.org/10.24908/ss.v2i2/3.3369

39. Edmond G (2013) Just truth? Carefully applying history, philosophy and sociology of science to the forensic use of CCTV images. Stud Hist Philos Sci Part C Stud Hist Philos Biol Biomed Sci 44:80–91. https://doi.org/10.1016/j.shpsc.2012.09.004

40. Bromby M (2003) At face value? New Law J Expert Witn Suppl 28:302–303

41. Thompson WC, Taroni F, Aitken CGG (2003) How the Probability of a False Positive Affects the Value of DNA Evidence. J Forensic Sci 48:2001171. https://doi.org/10.1520/jfs2001171

42. Dewhurst SA, Hay DC, Wickham LHV (2005) Distinctiveness, typicality, and recollective experience in face recognition: A principal components analysis. Psychon Bull Rev 12:1032–1037. https://doi.org/10.3758/BF03206439

43. Lee K, Byatt G, Rhodes G (2000) Caricature Effects, Distinctiveness, and Identification: Testing the Face-Space Framework. Psychol Sci 11:379–385. https://doi.org/10.1111/1467-9280.00274

44. Gibelli D, Obertová Z, Ritz-Timme S, et al (2016) The identification of living persons on images: A literature review. Leg Med 19:52–60. https://doi.org/10.1016/j.legalmed.2016.02.001

45. Damjanovski V (2014) CCTV from light to pixels, 3rd ed. Elsevier, Oxford

46. Ward D (2013) Testing Camera Height vs Image Quality. Pennsylvania, USA. Available at: https://ipvm.com/reports/testing-camera-height

47. Cohen N, Gattuso J, MacLennan-Brown K (2009) CCTV Operational Requirements Manual. Home Office Scientific Development Branch, Sandridge, UK. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/378443/28_09_CCTV_OR_Manual2835.pdf

48. Zhang X, Gao Y (2009) Face recognition across pose: A review. Pattern Recognit 42:2876–2896. https://doi.org/10.1016/j.patcog.2009.04.017

49. Burton A, Bruce V, Hancock PJB (1999) From pixels to people: A model of familiar face recognition. Cogn Sci 23:1–31. https://doi.org/10.1016/S0364-0213(99)80050-0

50. Bruce V (1994) Stability from Variation: The Case of Face Recognition The M.D. Vernon Memorial Lecture. Q J Exp Psychol Sect A 47:5–28. https://doi.org/10.1080/14640749408401141

51. Davis JP, Maigut A, Forrest C (2019) The wisdom of the crowd: A case of post- to ante-mortem face matching by police super-recognisers. Forensic Sci Int 302:109910.

https://doi.org/10.1016/j.forsciint.2019.109910

52. Roelofse MM, Steyn M, Becker PJ (2008) Photo identification: Facial metrical and morphological features in South African males. Forensic Sci Int 177:168–175. https://doi.org/10.1016/j.forsciint.2007.12.003

53. Vanezis P, Lu D, Cockburn J, et al (1996) Morphological Classification of Facial Features in Adult Caucasian Males Based on an Assessment of Photographs of 50 Subjects. J Forensic Sci 41:13998J. https://doi.org/10.1520/jfs13998j

54. Dodd V (2018) UK police use of facial recognition technology a failure, says report. Guard. Available at: https://www.theguardian.com/uk-news/2018/may/15/uk-police-use-of-facial-recognition-technology-failure

55. Press Association (2018) Welsh police wrongly identify thousands as potential criminals. Guard. Available at: https://www.theguardian.com/uk-news/2018/may/05/welsh-police-wrongly-identify-thousands-as-potential-criminals

56. Wall M (2019) Biased and wrong? Facial recognition tech in the dock. BBC News 1–12. Available at: https://www.bbc.com/news/business-48842750

57. Grother P, Ngan M, Hanaoka K (2019) Face recognition vendor test part 3: demographic effects. Gaithersburg, MD. https://doi.org/10.6028/NIST.IR.8280

58. Morrison R (2019) 'Racist' facial recognition technology used in law enforcement, banking and schools misidentifies African American and Asian people 100 times more often than whites, study shows. DailyMail Online. Available at: https://www.dailymail.co.uk/sciencetech/article-7811779/Massive-errors-facial-recognition-tech-US-study.html

59. Belhumeur PN, Jacobs DW, Kriegman DJ, Kumar N (2013) Localizing parts of faces using a consensus of exemplars. IEEE Trans Pattern Anal Mach Intell 35:2930–2940. https://doi.org/10.1109/TPAMI.2013.23

60. Burgos-Artizzu XP, Perona P, Dollar P (2013) Robust face landmark estimation under occlusion. Proc IEEE Int Conf Comput Vis 1513–1520. https://doi.org/10.1109/ICCV.2013.191

61. Sagonas C, Antonakos E, Tzimiropoulos G, et al (2015) 300 Faces In-The-Wild Challenge: database and results. Image Vis Comput 47:3–18.

https://doi.org/10.1016/j.imavis.2016.01.002

62. Ellis HD, Shepherd JW, Davies GM (1979) Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition. Perception 8:431–439. https://doi.org/10.1068/p080431

63. Megreya AM, Burton AM (2006) Unfamiliar faces are not faces evidence from a matching task. Mem Cognit 34:865–876. https://doi.org/10.3758/BF03193433

64. Bindemann M, Johnston RA (2017) Understanding how unfamiliar faces become familiar: Introduction to a special issue on face learning. Q J Exp Psychol 70:859–862. https://doi.org/10.1080/17470218.2016.1267235

65. Facial Identification Scientific Working Group (2012) FISWG Guidelines for Facial Comparison Methods. Available at: https://www.fiswg.org/FISWG_GuidelinesforFacialComparisonMethods_v1.0_2012_02_02.pdf

66. Alenezi HM, Bindemann M, Fysh MC, Johnston RA (2015) Face matching in a long task: Enforced rest and desk-switching cannot maintain identification accuracy. PeerJ 2015:3:e1184. https://doi.org/10.7717/peerj.1184

Tables

Table 1 Face pool trials by analysis: cohort subsample breakdown. Test trials were single normal trials including one target and nine foils and one correct match, No-match trials included one target and ten foil images. ST = standard, WT = wildtype, CCTV = Closed Circuit Television.

Analysis Cohort	Test Trials	No-match Trials	Repeat Trials	Analysis Trials	Total Trials
WT photograph to ST photographs	59	8	8	67	75
CCTV to ST photographs	86	8	6	94	100

Table 2 Facial pose distribution throughout face pool sample. ST = standard, WT = wildtype, CCTV = Closed Circuit Television

Analysis Cohort	Anterior View	Left 45° View	Left Lateral View	Right 45° View	Right Lateral View	Total Analysis Trials

WT photograph to ST photographs	14	14	13	13	13	67
CCTV standard to ST photographs	18	17	17	20	22	94

Table 3 Summarized list of facial components for facial comparison by morphological analysis [29].

Number	Facial Component
1	Skin (texture, colour, tone etc.)
2	Face/Head Outline (contour)
3	Face/Head Composition (proportions)
4	Hairline/Baldness Pattern (detailed shape)
5	Forehead (relative size and contour)
6	Eyebrows (angle, shape, relative size and hair details)
7	Eyes (colour, position, angle, relative size etc.)
8	Cheeks (prominence and dimples)
9	Nose (shape, position, symmetry, relative size etc.)
10	Ears (shape, position, protrusion and relative size)
11	Mouth (angulation, shape, symmetry, relative size etc.)
12	Chin/Jawline (shape, relative size, symmetry, prominence, and definition)
13	Neck (relative size, prominence, folds, veins etc.)
14	Facial Hair (presence, distribution, length, style, hair details etc.)
15	Facial Lines (creases, wrinkles etc.)
16	Scars (location, shape, orientation, relative size, tonality, depth etc.)
17	Facial Marks (moles, discoloration patches, freckles etc.)
18	Alterations (tattoos, piercings etc.)

Table 4 Contingency table reference for confusion matrix of positive and negative matches (adapted from Kuhn [36]). Formulae employed on each subset of analysis to compare predictive quality of analyses rely on this contingency table.

	True face match	*True face exclusion*
Matched by analyst	True Positive (TP)	False Positive (FP)

Excluded by analyst	False Negative (FN)	True Negative (TN)

Table 5 Summary of intra- and inter-observer agreements across all cohorts of analysis

	Intra-observer agreement	p-value	Inter-observer agreement	p-value
Photograph-to-photograph	0.934	< 0.001	0.921	= 0.001
CCTV to photograph	0.807	< 0.001	0.743	< 0.001

Table 6 Contingency table for confusion matrix of photograph-to-photograph comparison. TP = true positive; FP = false positive; TN = true negative; FN = false negative; IA = inconclusive analysis

	Actual face match	*Actual face exclusion*		*Actual face match*	*Actual face exclusion*
Matched by analyst	TP = 61	FP = 1	*Ideal perfect match*	TP = 62	FP = 0
Excluded by analyst	FN = 1	TN = 607	*Ideal perfect exclusions*	FN = 0	TN = 608

Table 7 Contingency table reference for confusion matrix of CCTV to photograph comparison**.** TP = true positive; FP = false positive; TN = true negative; FN = false negative; IA = inconclusive analysis

	Actual face match	*Actual face exclusion*		*Actual face match*	*Actual face exclusion*
Matched by analyst	TP = 76	FP = 43	*Ideal perfect match*	TP = 86	FP = 0
Excluded by analyst	FN = 10	TN = 785	*Ideal perfect exclusions*	FN = 0	TN = 828

Table 8 Confusion matrix function results and chance corrected accuracies across the lenient iteration of analysis

	Photograph-to-Photograph	*CCTV to Photograph*
Overall Raw Accuracy	0.997	0.942
95% Confidence Interval	0.989 – 1.000	0.925 - 0.956
No Information Rate (NIR)	0.908	0.906
P-value (Acc>NIR)	< 0.001	< 0.001
McNemar's Test P-value	1	< 0.001
Chance Corrected Accuracy (CCA)	98.2%	71.0%
P-value (for CCA)	< 0.001	< 0.001

Table 9 Tabulation of all predictive quality values calculated across the lenient iteration of analyses

	Photograph-to-Photograph	*CCTV to Photograph*
Sensitivity (Recall)	0.984	0.884
Specificity	0.998	0.948
Prevalence	0.093	0.094
Positive Predictive Value (PPV)	0.984	0.639
Negative Predictive Value (NPV)	0.998	0.987
Detection Rate (DR)	0.091	0.083
Detection Prevalence (DP)	0.093	0.130
Balanced Accuracy	0.991	0.916
Precision	0.984	0.639
F_1-Score	0.984	0.742

Table 10 Contingency table reference for confusion matrix of the strict iteration of the photograph-to-photograph comparisons analysed (left), compared to the ideal perfect matches and exclusions for direct comparison of performance of the facial comparisons (right). TP = true positive; FP = false positive; TN = true negative; FN = false negative; IA = inconclusive analysis

	Actual face match	*Actual face exclusion*		*Actual face match*	*Actual face exclusion*
Matched by analyst	TP = 61	FP = 0	*Ideal perfect match*	TP = 70	FP = 0

Excluded by analyst	FN = 1	TN = 608	*Ideal perfect exclusions*	FN = 0	TN = 608

Table 11 Contingency table reference for confusion matrix of the strict iteration of the CCTV to photograph comparisons analysed (left), compared to the ideal perfect matches and exclusions for direct comparison of performance of the facial comparisons (right). TP = true positive; FP = false positive; TN = true negative; FN = false negative; IA = inconclusive analysis

	Actual face match	*Actual face exclusion*		*Actual face match*	*Actual face exclusion*
Matched by analyst	TP = 66	FP = 5	*Ideal perfect match*	TP = 86	FP = 0
Excluded by analyst	FN = 20	TN = 823	*Ideal perfect exclusions*	FN = 0	TN = 828

Table 12 Confusion matrix function results across the strict iteration of each cohort of analysis

	Photograph-to-Photograph	*CCTV to Photograph*
Overall Raw Accuracy	0.999	0.973
95% Confidence Interval	0.992 – 1.000	0.960 - 0.982
No Information Rate (NIR)	0.908	0.906
P-value (Acc>NIR)	< 0.001	< 0.001
McNemar's Test P-value	1	0.005
Chance Corrected Accuracy (CCA)	99.1%	82.6%
P-value (for CCA)	< 0.001	< 0.001

Table 13 Tabulation of the predictive quality values calculated across the strict iteration of each cohort of analysis

	Photograph-to-Photograph	*CCTV to Photograph*
Sensitivity (Recall)	0.984	0.767
Specificity	1.000	0.994
Prevalence	0.093	0.094
Positive Predictive Value (PPV)	1.000	0.930

Negative Predictive Value (NPV)	0.998	0.977
Detection Rate (DR)	0.091	0.072
Detection Prevalence (DP)	0.091	0.078
Balanced Accuracy	0.992	0.881
Precision	1.000	0.930
F_1-Score	0.992	0.841

Table 14 False positive rates (FPR) and false negative rates (FNR) calculated across both lenient and the strict iterations of the analysis cohorts

	Photograph-to-photograph	CCTV to photograph
FPR Lenient iteration	0.16%	5.19%
FPR Strict iteration	0%	0.60%
FNR Lenient iteration	1.61%	11.63%
FNR Strict iteration	1.61%	23.26%

Figures

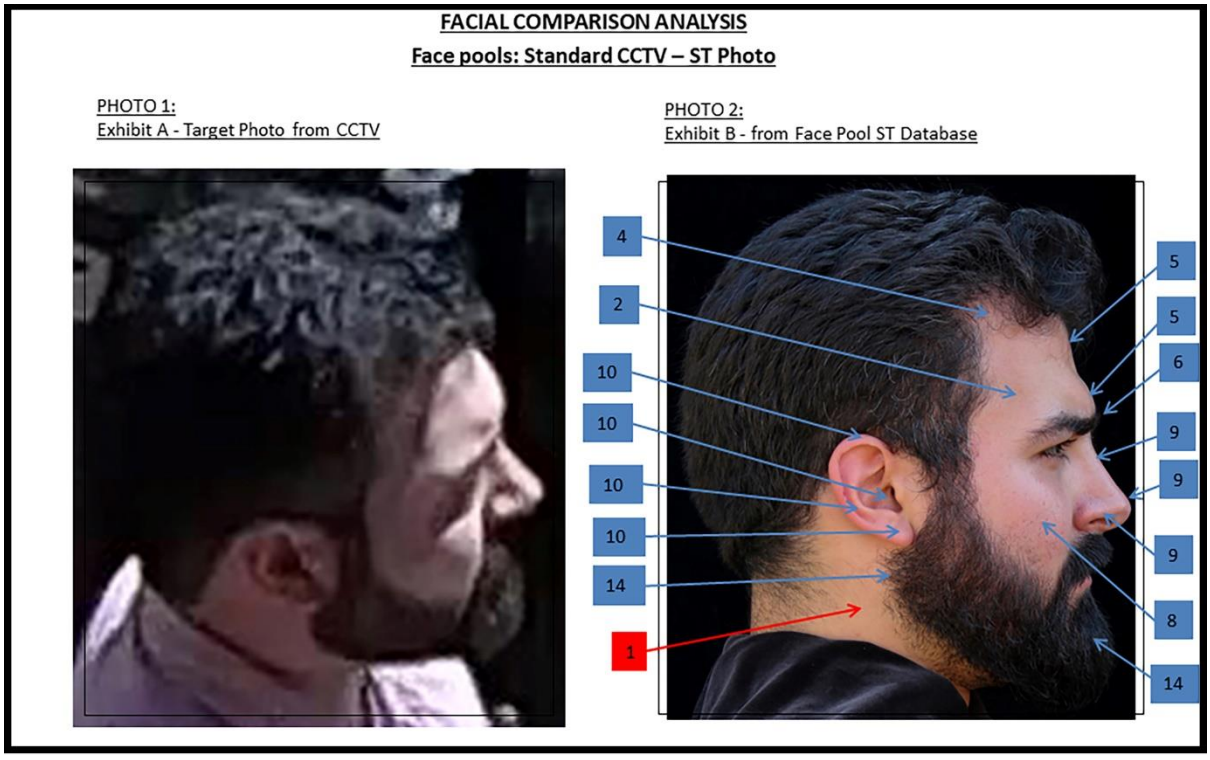


Fig. 1 Facial comparison analysis example. A target image from CCTV or wildtype (WT) photograph was placed next to each possible match based on possible positive matching during holistic analysis. Standardised (ST) photograph on the right was compared to target CCTV

extracted image on the left using the FISWG facial component list [29]. Similar (blue) and dissimilar (red) facial components were highlighted by facial component number



Fig. 2 Bar graphs of the facial comparison outcomes under a lenient (A) and strict (B) iteration within the photograph-to-photograph cohort by trial. TP = true positive; FP = false positive; TN = true negative; FN = false negative; IA = inconclusive analysis



Fig. 3 Bar graphs of the facial comparison outcomes under a lenient (A) and strict (B) iteration within the CCTV to photograph cohort by trial. TP = true positive; FP = false positive; TN = true negative; FN = false negative; IA = inconclusive analysis

4. Performance of forensic facial comparison by morphological analysis across optimal and suboptimal CCTV settings

Nicholas Bacci, Maryna Steyn, Nanette Briers

Science & Justice, 2021 (Submitted and under review)

Performance of forensic facial comparison by morphological analysis across optimal and suboptimal CCTV settings

Nicholas Bacci[a], Maryna Steyn[a], Nanette Briers[a]

[a]Human Variation and Identification Research Unit, School of Anatomical Sciences, University of the Witwatersrand, Johannesburg, 7 York Road, Parktown, 2193, South Africa

Email addresses:

Nicholas Bacci: nicholasbacci@gmail.com

Maryna Steyn: maryna.steyn@wits.ac.za

Nanette Briers: nanette.briers@wits.ac.za

Corresponding author:

Mr Nicholas Bacci, MSc

Present address: Human Variation and Identification Research Unit, School of Anatomical Sciences, University of the Witwatersrand, 7 York Road, Parktown, 2193, South Africa

Email: nicholasbacci@gmail.com and nicholas.bacci@wits.ac.za

Tel: +27-11-717-2078

ORCID numbers:

Mr Nicholas Bacci: 0000-0001-6858-7598
Dr Nanette Briers: 0000-0002-8544-6049
Prof Maryna Steyn: 0000-0002-0215-9723

Declarations:

Funding The current study was conducted with support from the South African National Research Foundation (NRF) and the J.J.J. Smieszeck Fellowship from the School of Anatomical Sciences, University of the Witwatersrand (DAAD-NRF and J.J.J. Smieszeck Fellowship funds awarded to N. Bacci (Grant No.: 11858) and NRF funds awarded to N. Briers as part of the Improving Methodologies and Practices in Craniofacial Identification (Grant No.: CSUR160425163022; UID: 106031). Any opinions, findings and conclusions or recommendations expressed in this study are those of the authors and therefore the NRF and

University of the Witwatersrand, Johannesburg, do not accept any liability in regard thereto. The funders had no involvement in study design, data collection, analysis, or interpretation, writing of the manuscript or submission.

Conflicts of interest/Competing interests: All authors affirm that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Ethics approval: Approval for the present study was obtained from the Human Research Ethics Committee (Medical) of the University of the Witwatersrand, Johannesburg. Ethics clearance certificate number: M171026.

Availability of data and material: The data analysed during the current study are not publicly available due the sensitive nature of the Wits Face Database (http://dx.doi.org.10.18772/10539/20690). A sample of the database is available at https://hdl.handle.net/10539/29924. Restrictions apply to accessing the database or part thereof, which was used under license for the current study. Access is limited to formal request and conditional approval by the School of Anatomical Sciences Collections Committee, strictly for non-commercial research.

Acknowledgements

Acknowledgement is owed to all the participants who agreed to be photographed and recorded for the development of the Wits Face Database from which the face pool samples used in the current were composed. Thanks are due to Joshua Davimes for his instrumental work in the development of the Wits Face Database. Tamara Lottering is also acknowledged for her assistance in composing the face pools as is Gideon LeRoux for his assistance in the capturing and extracting of the CCTV recordings from the University security systems. Gratitude is also owed to Tobias Houlton for his aid in study validation.

Thanks are also due to the volunteers who aided in participant recruitment: Jesse Fredericks, Kiveshen Pillay, Rethabile Masiu, Sameerah Sallie, Daniel Munesamy, Laurette Joubert, Jordan Swiegers, Betty Mkabela, Johannes P. Meyer, Amy Spies, Natasha Loubser, Nicole Virgili, Dan-Joel Lukumbi, Tamara Lottering, Mathabatha Ntjie, Claudia Landsman, Raheema Dalika, Merete Goosen, Stephanie Souris, Rabelani Negota, Mahlatse Mahasha, Jessica Manavhela.

Novelty statement

The present study is the first to test forensic facial comparison by morphological analysis using the Facial Identification Scientific Working Group feature list in both a high-resolution digital CCTV and low-resolution analogue CCTV samples. In so doing it was possible to determine the severity of decrease in performance that suboptimal CCTV footage has on identification, while also showing the benefits of using a feature list in a poor-quality sample in the form a low false positive rate. It is also highlighted that relying on hit rates for studies of this nature can lead to severely flawed conclusions overestimating performance of identification methods.

Highlights

- Morphological analysis has not yet been tested under different CCTV conditions.
- Excellent accuracies were found in high-resolution digital CCTV images.
- Low-resolution analogue CCTV data negatively impacted identification capability.
- Low false positive rates were found for both optimal and suboptimal CCTV data.
- Chance corrected accuracy is recommended to evaluate facial comparison outcomes.

Abstract

Facial comparison is an important yet understudied discipline in forensics. The recommended method for facial comparison in a forensic setting involves morphological analysis (MA) with the use of a facial feature list. The performance of this approach has not been tested across various closed-circuit television (CCTV) conditions. This is of particular concern as video and image data available to the police is often varied and of subpar conditions. The present study aimed at testing MA across two types of CCTV data, representing ideal and less than ideal settings. The study was conducted on a subset of the Wits Face Database arranged in a total of 225 face pools. Each face pool consisted of a target image obtained from either a high-definition digital CCTV camera or a low-definition analogue CCTV camera in monochrome, contrasted to 10 possible matches. The face pools were analysed and scored using MA and confusion matrices were used to analyse the outcomes. A notably high chance corrected accuracy (CCA) (97.3%) and reliability (0.969) was identified across the digital CCTV sample, while in the analogue CCTV sample MA appeared to underperform both in accuracy (CCA: 33.1%) and reliability (0.529). The majority of the errors in scoring resulted in false negatives in the analogue sample (75.2%), while across both CCTV conditions false positives were low

(digital: 0.3%; analogue: 1.2%). Even though hit rates appeared deceptively high in the analogue sample, the various measures of performance used and particularly the chance corrected accuracy highlighted its shortfalls. Overall, CCTV recording quality appears closely associated to MA performance, despite the favourable error rates when using the Facial Identification Scientific Working Group feature list.

Keywords: forensic facial comparison, morphological analysis, CCTV, analogue, digital, facial identification.

1.1 Introduction

As eyewitness identification has since the 1970s been considered unreliable and surveillance evidence has become increasingly more common [1], facial identification from photographic and closed-circuit television (CCTV) surveillance evidence has become common practice [2], particularly conducted in courts by juries [3] or experts [4]. CCTV data availability has increased due to the escalation in surveillance system installations throughout the world [5]. Facial identification (ID) is, as agreed upon by many authors, a challenging task [6–8] due to human variation, facial expression, ageing and other biological factors [e.g. 7–10]. This difficulty is further compounded by poor image quality, which effectively results in a decreased amount of data for a facial comparison expert to base their conclusions on. The effect on forensic facial comparison of poor-quality data is highly problematic in a legal context where its application can have a most damaging impact if misused or misinterpreted.

Facial ID in the legal system is carried out by forensic facial comparison using morphological analysis [13,14]. Studies that compare the performance of facial comparison generally utilise highly standardised images with optimal conditions and even so result in error rates ranging between 10 and 30% [8,15]. While few studies have investigated the performance of forensic facial comparison [4,16–18], even fewer have considered the effects of low-quality images or recordings [19,20]. A recent study has found morphological analysis, when using a facial feature list, to be highly reliable and accurate both in photographic and CCTV images [18], however, it did not investigate low-quality CCTV data. Unfortunately, the majority of images analysed in a forensic setting tend to be of low to medium quality [21]. These low-quality images can originate from subpar, outdated equipment or poor recording conditions.

Most recent systems use Internet Protocol (IP) cameras with relatively high resolutions, although many outdated installations may still be based on analogue Digital Video Recording (DVR) systems [22]. Analogue CCTV systems have low resolution recordings and older

systems record in monochrome, compounding the information loss even further [23]. Modern CCTV systems have higher image resolutions with full colour recordings that provide more information for an analyst and are often connected to computer enhanced surveillance networks [24]. Analogue systems have dominated the CCTV markets both in South Africa and globally up to recent years, likely due to the low costs and accessibility of these systems [25,26]. A new increase in IP camera installations begun in 2015 [27,28], however, despite the multiple benefits of IP surveillance systems many consumers have resisted making the shift, likely as a result of finding analogue systems trustworthy and satisfactory [29].

Specifications vary between analogue and IP cameras that can limit the quality of the footage obtained, yet perhaps the most defining one is resolution. Analogue camera resolutions are measured in Television Lines (TVL) and do not directly correspond to pixel resolutions, commonly used today in the digital media era [22]. Resolutions of analogue cameras tend to vary from 480 to 1000TVL with cameras of 700TVL or more considered high definition [22]. A typical 700TVL analogue camera would result in a practical maximum resolution of 934 x 560 pixels (based on Damjanovski [22]). The resulting images would hence include much lower amounts of detail and provide a pixelated or blurry appearance on high resolution monitors.

Modern high definition (HD) IP CCTV expresses resolution in pixels directly and is usually presented in megapixels (MP). A typical 4 MP IP camera has a resolution of 2560 x 1440 pixels, almost three orders of magnitude greater than a typical 700TVL analogue camera. The greatest concerns for image quality with HD cameras, though, are compression artefacts [22]. Compression used in CCTV systems is considered lossy by nature, as quality loss occurs in the process of compression; once performed in the camera it cannot be removed and may result in noticeable distortion of recordings. This distortion usually is seen as pixelation, basis patterns, blurring, ringing, and other less common forms of artefacts [22,30]. These compression artefacts can severely hinder and inhibit interpretation and analysis [30,31]. While recent strides has been made that could drastically improve performance of facial recognition systems [32], for most systems lossy compression artefacts remain a concern.

While many studies have looked broadly at facial identification [e.g. 8,32,33], the effects of poor quality and suboptimal condition CCTV images on facial identification have only been tested by a handful of authors [19,20,23,34]. Testing the application of facial identification across CCTV footage of various quality is crucial to demonstrate whether low quality

recordings can be deemed acceptable evidence in the legal system. Therefore, this study aimed at comparing the outcomes of facial comparison of high quality and low-quality CCTV recordings, using the Facial Identification Scientific Working Group (FISWG) list of morphological features commonly used for forensic facial comparison [35].

1.2 Materials and Methods

Recordings and photographs used for this study were drawn from the Wits Face Database and were arranged in face pools [36,37]. The participants recruited into the Wits Face Database were all informed, consenting adults and the present study was approved in full by the Human Research Ethics Committee (Medical) (clearance certificate number: M171026). The face pool sample included facial photographs and images extracted from CCTV recordings of a total of 219 consenting, known young adult male participants of South African descent. Each face pool consisted of a target snapshot of CCTV recordings and 10 standardised database photographs. Standardised database photographs (ST) were captured at a height of 1600 mm using a Canon 1300D 18MP DSLR camera (18-55 mm DC Canon lens) at an objective to face distance of 1500 mm using a black velvet cloth backdrop and cloth to obscure the participants' clothing. Captured photographs had an image resolution of 3456 x 4320 pixels at 300 dpi and were obtained at five facial views: anterior, right 45-degree, right lateral, left 45-degree, and left lateral view.

CCTV recordings were obtained through two different surveillance systems, a low resolution Securi-Prod 1/3" Sony Effio E 700TVL (NTSC signal system) indoor dome analogue camera (model: CC217, 2.8 – 12 mm vari-focal lens) positioned at ceiling height (2500mm) and a high-definition IP camera (HikVision, model: DS-2CD2142FWD-I, 4-megapixel, maximum resolution of 2688 x 1520, 4 mm fixed lens, aperture F2.0) installed at "eye-level" (1700mm). These surveillance systems were not specifically installed for the purposes of database development or the present study - existing systems were used. The effective pixel resolution of the analogue camera was approximately 976 x 494, which is roughly the equivalent of 0.5 of a megapixel. The objective to face distances of the eye-level digital and analogue cameras were 800 mm and 2920 mm, respectively. The angle of incidence between the face and the objective of the camera was effectively zero at the digital camera and approximately 22° under the analogue camera. A detailed description of the Wits Face Database image capture process have been described at length [37]. These images and photographs were combined into face

pool trials based on broadly visually similar individuals to test the performance of the FISWG facial feature list for individual identification under different actualistic conditions.

A total of 225 face pool trials were used, each composed of a single target facial photograph grouped with either nine foil images and one correctly matching image (test trials) or 10 foil images (no-match trials). The foil images were assigned to face pools based on relative facial resemblance and apparent age by a research assistant, to allow blinded analysis. Face pools were grouped by the target type of facial images and arranged in corresponding views, or facial poses. Digital CCTV face pools (n=95) included a captured high-resolution frame as a target image, to be compared to 10 facial images from the database composed of standardised (ST) optimal photographs with clothing obscured. The selected frame was the most ideal potential image from the footage available, which may not necessarily reflect realistic situations where multiple still images can be extracted for comparison, however, a level of standardisation was required for experimentation. Similarly, the analogue CCTV face pools (n=130) were composed of a captured frame from the analogue CCTV recordings and 10 ST facial images. The images of the analogue cohort were also converted into greyscale images using Photoshop (CS 6) to simulate monochrome low-resolution cameras, as these conditions were considered the least optimal possible for CCTV images. The CCTV extracted images were cropped similarly to the ST photographs, with the face representing the central $1/9^{th}$ of the image and at an aspect ratio of 4:5. The resulting images had an approximate resolution of 2199 x 2749 pixels at the native 72 dpi of the digital CCTV recordings and 64 x 80 pixels at the native 72 dpi of the extracted analogue recordings. The face was effectively composed of approximately 800 x 1000 pixels for the digital and 18 x 26 pixels for the analogue cameras, respectively. A maximum of 15 repeat uses of ST photograph was permitted in the composition of the face pools to prevent familiarisation of the principal analyst.

To ensure that observations were reliably performed and repeatable, repeat trials were included throughout face pools to allow for intra-observer error calculation. These repeat trails were specifically assigned positions within the face pool sample at isolated slots unknown to the analyst to prevent familiarisation. A detailed rundown of the sample for each analysis cohort is found in Table 1.

Relatively even distribution of facial poses, or views, were assigned throughout each cohort (Table 2), while still retaining correspondence of views in each individual face pool. The use of the various facial poses was specifically included to permit a broad review facial features

across the entire face without giving preference to certain features by only focusing on a singular view of the face.

The FISWG guidelines on facial comparison were utilised for the analysis process. To replicate actualistic conditions, potential matches were briefly examined using holistic review according to FISWG guidelines [14,38]. The holistic review process was performed by reviewing the entirety of the face under a controlled time limit of five seconds to avoid detailed comparison of specific facial components. During this review, a face was deemed either as a possible match (1), not a match (2), or no analysis possible (3). Any possible matches were then further analysed through morphological analysis. Morphological analysis (MA) comparison was conducted using the FISWG facial component feature and descriptor list [35].

For morphological analysis, the target image was placed next to the ST image in Microsoft PowerPoint (365, v. 16). The images were then adjusted to resemble the orientation and size of one another while maintaining the original aspect ratio. Morphological features of the face were then compared and contrasted for similarities or dissimilarities (Figures 1 and 2). These were identified by their numerical value on the FISWG facial component list and visually assessed to reach a decision for each possible match. The final decisions for each facial comparison were obtained through the use of an adapted version of the levels of agreement system used by the South African Police Services [13], as previously done in a study validating the FISWG feature list for morphological analysis [18].

In order to calculate hit rate, the conclusions were grouped as either being a correct match, not being a match, or an inconclusive analysis. These conclusions, inclusive of the trials excluded as “not matching” in the holistic review phase of the analysis, were then compared to the true matchup of the images to calculate the hit rate. Resulting from this comparison, the matchup outcomes were marked as a true positive (TP), a false positive (FP), a false negative (FN), a true negative (TN), or an inconclusive analysis (IA). Following the same approach used in another study testing the FISWG facial feature list [18], the level of agreement scores were grouped for statistical analysis in two iterations - lenient and strict. The lenient iteration included the two scores closest to a positive match (scores 1 and 2 on the level of agreement scale) as a match, the two scores closest to a negative match (scores of 3 and 4) as not matches and the last score was an inconclusive analysis which was excluded from further analyses. The strict iteration was aimed at reducing expected false positives when scoring matches as having some level of similarity in visible morphological features (score 2). As such, positive matches

were considered only comparisons scored as 1 while scores of 2, 3 and 4 were all grouped as negative matches.

Statistical analysis was conducted with the R Studio statistics programme (version 3.6.3) [39]. Multiple types of accuracies were calculated for facial identification in each cohort, including raw overall accuracy, balanced accuracy, and chance corrected accuracy by conducting a weighted Cohen's kappa. Target accuracy for successful performance was taken as a kappa value of 0.80 or above [40]. Intra-observer agreement was also calculated using weighted Cohen's kappa.

Using the matchup outcomes, a complete confusion matrix including multiple measures of predictive power were calculated from the resulting contingency table (Table 3). A raw overall accuracy, or hit rate, was calculated using the confusion matrix function in R, by dividing the total correct scores by the total sample, and a 95% confidence interval was also calculated from the contingency table. This hit rate accuracy was then compared to a No Information Rate (NIR), defined as the largest proportion of the observed matchup outcomes [41]. A p-value was calculated from this comparison, to confirm whether the accuracy was statistically higher than the NIR as an indicator of a good predictive model [42]. To determine whether the scored matches have a different proportion of errors (between false positive and false negative outcomes), a McNemar's Test was calculated as part of the function [43]. A p-value below $\alpha = 0.05$ indicated a significant difference in the proportion of errors within the sample. This comparison effectively indicated whether the analysis method errs relatively evenly in false positives and false negatives. False positive and false negative rates were also calculated to objectively review the performance of MA in methods.

To calculate the quality of predictive power for morphological analysis, the formulae included in the confusion matrix function and drawn from the Caret R package [41,42] are defined below and shown in Table 4. Sensitivity, or true positive rate or recall, was defined as the measure of how effectively a test can perform correct matches (Table 4). Specificity, or true negative rate, is a measure of how effective a test can correctly discard a match (Table 4). Prevalence was the total number of truly correct matches in the entire cohort sample (Table 4). The positive predictive value (PPV), also known as precision, indicated the probability of a trial scored as a match being truly a correct match (Table 4) and negative predictive value (NPV) was the probability of a trial scored as an exclusion being truly a correct exclusion (Table 4). Detection rate was the proportion of the cohort sample correctly matched (Table 4) and detection

prevalence was the proportion of matches assigned by the analyst (Table 4). Under perfect circumstances, the detection rate and prevalence would be identical or as close to the prevalence value as possible, which was in the range of 9 to 10% in this study due to the organisation of the sample. An adjusted accuracy calculation that considers the unbalanced distribution of the sample, was the balanced accuracy calculated using equation 8 making use of sensitivity and specificity. Lastly, the F-Score is a weighted (harmonic) mean of precision (or PPV) and sensitivity which takes into consideration both false positive and false negative outcomes (Table 4). F-Scores can be weighted according to the requirements of the study by controlling the weight of precision and sensitivity in the equation. In the present study equal weighting was given to both variables so that β was assigned a value of 1 making the statistic an F_1-Score (Table 4).

1.3 Results

Intra-observer agreement was found highly consistent in the digital CCTV sample with an almost perfect level of agreement ($\kappa = 0.969$; $p < 0.001$). However, within the analogue CCTV sample, a drastically lower level of agreement was found with a moderate result ($\kappa = 0.529$; < 0.001) (Table 5). Similarly, while inter-observer agreement in the digital sample was high ($\kappa = 0.773$; $p < 0.001$) a sharp decrease ($\kappa = 0.378$; $p < 0.001$) was noted in the analogue sample (Table 5).

The overall results from the individual face pool facial comparisons are presented in Figures 3 and 4, where the occurrence of true positives, false positives, false negatives, true negatives, and incomplete analyses are shown, as assigned during the analyses conducted in each trial. Any trials classified as inconclusive analyses (totalling 99 individual comparisons all within the analogue CCTV cohort) and repeat trials (8 face pools in each cohort of analysis, each including 10 facial comparison trials) were not included in the remaining results presented. The large number of inconclusive analyses is an indication of the drastic decrease in quality of the footage obtained by the analogue CCTV camera compared to the good quality of images obtained from the digital camera.

Morphological analysis, under a lenient iteration, performed particularly well in the digital sample as seen from the confusion matrices calculated (Tables 6 and 7). This was evident in the perfect true positive rate (100%) and very high true negative rate (98.9%), paired with a low false positive rate (1.1%), of the digital CCTV sample (Table 6). However, much poorer results were seen in the analogue sample with lower true positive rate (56.2%) and true negative

rate (95.4%) and notably higher error rates, with a small increase in false positives (4.6%) and particularly in relation to false negatives (43.8%) (Table 7). Raw accuracies (RA) suggested that the outcomes are acceptable in both samples with a high performance in both the digital (0.990, CI = 0.981 - 0.995) and analogue (0.918, CI = 0.900 - 0.933) CCTV recordings (Table 8). However, this apparent optimal performance of MA across both samples was more indicative of the shortcomings of relying on a simple hit rate to assess the performance of a specific method as the composition of the sample and chance are not taken into consideration. The inclusion of the no information rate (NIR) comparison, however, highlights the poorer performance of MA in the analogue sample as the RA and the confidence interval are relatively close to the NIR of the sample, resulting in a p value above $\alpha = 0.05$ (Table 8). However, the digital sample performed well over the NIR value and suggests a good outcome in the high-resolution sample.

A much greater disparity in performance of MA across the samples is demonstrated by the chance corrected accuracies (CCA), with a particularly high accuracy in the digital sample ($\kappa = 0.942$; $p < 0.001$) and low accuracy in the analogue sample ($\kappa = 0.514$; $p < 0.001$). A difference in the proportions of errors was found in the digital sample (McNemar's test, Table 8), as there was a complete lack of false negatives in the sample (Table 6).

However, this same relative proportion of errors was not seen in the analogue sample (McNemar's test, Table 8), where MA appeared to result in both FP and FN outcomes almost evenly (Table 7). Sensitivity was perfect for the digital cohort and much lower (0.562) for the analogue sample (Table 9), as almost half the trials were incorrectly assigned as false negatives (Table 7). However, specificity appears to be similar in both samples, most likely due to the percentage composition of the sample, as the majority of the sample consists of negative matches by design, as shown by the identical prevalence across samples (Table 9).

Within the digital cohort, the positive predictive value (PPV) was quite high at 0.900 (Table 9), which indicates a strong ability of MA in identifying TP and avoiding FP outcomes at the same time. By contrast, in the analogue sample, PPV was much lower at 0.557 (Table 9), as suggested by the many false positives resulting in that analysis. Negative predictive value is perfect for digital and high for analogue CCTV footage (0.955), however, the sample

composition, which is heavily skewed towards TN outcomes, should be taken into consideration.

The detection rate of the digital sample (9.3%) approaches the detection prevalence (10.4%), indicative of an almost complete identification of correct matches in that sample, while the notably lower detection rate (5.2%), despite a similar detection prevalence (9.4%) seen in analogue footage demonstrates the difficulty in identifying matches overall (Table 9). Detection prevalence was close to the composition of the sample expectation in both samples, yet lower in the analogue sample, possibly due to the decrease of confidence in assigning a match due to the decreased quality.

An extremely high balanced accuracy (99.4%) was identified in the digital sample, while a decreased balanced accuracy in the analogue sample (75.8%) likely results from the low quality of footage (Table 9). The drastically lower precision in the analogue trials (55.7%) can be explained by the increased false positives in that cohort, while digital trials had a high precision of 90% (Table 9). F_1-Scores demonstrated a similar discrepancy with a much better performance of MA in the digital sample than the analogue sample (Table 9), as expected based on the high rates of false positives and negatives in the analogue sample (Table 7).

The strict iteration demonstrated a similar performance of MA in the digital sample with a decrease in the number of false positives. Conversely, in the analogue sample, the strict iteration resulted in a drastic increase in false negatives and a decrease in false positives (Table 11). This was clearly seen when comparing the false positive and false negative rates between the lenient iteration and the strict iteration, with fourfold decrease in false positives and a 75.2% rate of false negatives in the analogue sample (Table 11). The digital sample had a decrease of false positive rate to 0.3% with an increase in false negatives from 0% to a low value of 2.5% (Table 10).

This shift in false positives and negatives is accompanied by a marginal increase in hit rate accuracy in both samples in the strict iteration (Table 12). The increased hit rate accuracy of the analogue sample, resulting from the increased number of correct exclusions matched, was still below acceptable (0.920, CI = 0.902 – 0.935) as it was not significantly higher than the NIR (Table 12). Interestingly, while the proportion of errors was equal in the digital sample, across the analogue sample, it shifted from even to heavily weighted towards false negatives in this iteration as indicated by the McNemar's Test results (Table 12). Chance corrected

accuracy increased in the strict iteration of the digital sample (97.3%) and dramatically decreased in the analogue sample (33.1%) (Table 12).

With the strict iteration a decrease in sensitivity was also seen across both samples. This was marginal in the digital cohort and much more drastic in the analogue samples, while both samples saw a slight increase in specificity (Table 13). Similarly, PPV increased in both samples and NPV decreased in both samples (Table 13). However, PPV for the analogue sample still remained at a moderate level, while NPV remained in a high range. This suggests that even in considerably suboptimal footage a strict iteration remains of high value in negative prediction.

Detection rate for the digital sample remained in a similar range in the strict iteration, approaching 10% in dictating the high amount of true positive matches, while in the analogue sample the detection rate is halved in the strict iteration (Table 13), which corresponds with the much-increased rate of false negatives (Tables 10-11). Detection prevalence also decreased in both samples, expectedly with the decrease in what was considered matched by the analyst, however, it is important to note that the detection prevalence in the analogue sample was drastically lower.

A decrease in balanced accuracy is also noted in the strict iteration across both samples (digital: 98.6%, analogue: 0.618), with a greater decrease in the analogue sample (Table 14). Precision, however, appears to increase moderately from 0.900 to 0.975 in the digital CCTV sample and from 0.557 to 0.684 in the analogue CCTV sample, suggestive of a better categorisation system in the strict iteration. Lastly, the F_1-Score increased minimally from 0.947 to 0.975 in the digital cohort and decreased notably (from 0.559 to 0.364) in the analogue sample, indicative of the overall poorer performance of MA in the analogue sample.

1.4 Discussion

Widespread installations of surveillance systems [5] have increased the amount of image-based data available to the judicial system [2]. Until recently, CCTV installations were primarily analogue cameras [25]. While a new shift towards higher quality IP cameras has begun in the last five or more years [27], the majority of evidence available to forensic analysts is still of low to medium quality [21], making it important to test the efficacy of facial ID methodologies, especially morphological analysis, across various image qualities.

In the current study, the outcomes of morphological analysis [44], was tested in a high quality and optimally captured IP CCTV sample as well as a low quality, monochrome and sub-optimally captured analogue CCTV sample. This is the first study to test the accuracy of morphological analysis, and particularly the FISWG feature list, across different CCTV conditions. Morphological analysis (MA) has been found to be accurate and reliable in both photographic and CCTV recording data, with a better performance in more optimal high-quality photographic data over CCTV data [18]. Expectedly, in the present study MA was seen to be highly effective at identifying faces in the digital sample, with accuracies ranging from 94.2% (chance corrected) to 99% (hit rate and balanced accuracies) in the lenient iteration, and high levels of both intra- and inter-rater agreement (0.969 and 0.773 respectively). It also expectedly performed worse in the lower-quality analogue CCTV sample. Despite this lower performance, hit rate and balanced accuracies (91.8% and 75.8% respectively) in the analogue sample were high. These deceptively high accuracies were placed into perspective when a hit rate accuracy not statistically lower than the no information rate of the sample was found. This finding suggested an overall poor performance in the analogue sample that would not have been obvious by relying solely on hit rate accuracy. In fact, a low chance corrected accuracy of 51.4% and moderate to poor intra- and inter-rater agreement (0.529 and 0.378 respectively) were also found, indicating that chance corrected accuracy may be a much better indicator of true performance.

These results suggest that image conditions and quality influence both the reliability and accuracy of facial comparison as a whole, and MA specifically. This warrants reconsideration on the use of poor condition and quality image data for forensic facial comparison, as its reliability is low and the accuracy not good enough for application in the judicial context. In addition, this low accuracy and low reliability in the analogue sample's lenient iteration matches incorrectly resulted in both a high false positive rate (4.6%) and a considerably high false negative rate (43.8%), while in the digital sample few false positive outcomes (1.1%) resulted from the incorrect matches with no false negatives whatsoever.

When adjusting the cut-off scale for match assignment through the strict iteration of the data analysis, the performance of MA in digital comparisons increased minimally (chance corrected accuracy (CCA): 97.3%), while in the analogue sample a drastic decrease in performance was seen (CCA: 33.1%). This was noted despite an increase in hit rate accuracy across both samples. In conjunction, within the analogue sample, false positives decreased to levels seen in the digital sample's lenient iteration (1.2%), while false negatives almost doubled in

prevalence (75.2%). In the strict iteration of the digital sample, however, a decrease in false positive rate (0.3%) and some increase in false negatives resulted (2.5%). This suggested that in both good and low-quality images, a strict iteration mitigates the false positive rate at the cost of an increased false negative rate, with a much stronger effect experienced in the low-quality image sample. It is also important to note that almost 100 facial comparisons were not possible at all due to the low-quality of the analogue footage. Effectively meaning that almost 10% of the sample was impossible to analyse, which is quite worrisome.

When compared to the outcomes of another study that tested the FISWG feature list, also using the Wits Face Database [18], the present study MA performed notably better (97.3%) in the digital CCTV recordings obtained at eye-level than in traditional standard CCTV recordings (82.6%) [18], and only marginally worse than the photograph only sample (99.1%) [18]. This further reinforces the exceptionally good performance of MA with the aid of a systematic facial feature list in optimal and high-quality images. However, based on the current study's analogue CCTV results, when a low-quality sample is under consideration, MA underperforms, even with the use of a feature list. In addition, when compared to the present study, a study using a standard CCTV sample identified 26 truly inconclusive analyses [18], further demonstrating the discrepancy in quality and ability to effectively analyse poor quality CCTV recordings.

Other studies have also, as expected, found reduced accuracy under poor conditions [e.g. 18,22,45]. The unsurprising inferior performance of MA in the analogue sample is, however, difficult to compare to results from other studies. This difficulty arises from only a handful studies investigating the efficacy of morphological analysis, as well as the many variations of the experimental approach used for facial identification [46,47] and inconsistent reporting [16]. The majority of studies in facial identification compare the ability of multiple independent observers to broadly identify or match individuals in images [e.g. 4,45]. Irrespective of whether they follow a specific methodological approach (such as MA with a specific feature list), many studies that report on facial identification performance in video recordings fail to describe the exact quality of the recordings or report only hit rate accuracies [e.g. 17,18]. Based on the current study's results, it is clear that hit rates can actually be deceivingly high as they do not take chance into consideration. Qualitatively, the lowest quality recordings are used in a study by Lee *et al.* [19], where a hit rate of approximately 67% is reported, a hit rate lower than that identified in the present study even in the worst conditions. Other similar studies with slightly better quality video recordings, resulted in hit rate accuracies in similar ranges of 70 to 75% [17,45]. The hit rate presented for experts by Wilkinson and Evans [4], is the highest reported

from CCTV recordings (81%), with recordings of seemingly better quality than the analogue sample used in the present study. In the present study's analogue sample, the comparative hit rate is much higher (91.8%), likely due to the make-up of the sample including a greater number of true negative than true positive matches. However, the balanced accuracy (corrected for specificity and sensitivity), both in the lenient (75.8%) and strict iterations (61.8%), was in similar ranges to those in other studies.

When comparing false acceptance (false positive) and false rejection (false negative) rates, the present study had a much higher rate of false rejection in both lenient and strict iterations (43.8% and 75.2% respectively) in the poor-quality sample than what was reported by Lee and colleagues (10.5%) [19]. In addition, Lee *et al.,* [19] reported a much higher false acceptance rate (22.4%), more than twice their false rejection rate. In better video quality conditions, Wilkinson and Evans [4] identified a lower false rejection rate (8%) and a higher false acceptance (17%) rate. The overall lower error rates seen in the digital sample is further reinforcing the better performance of MA in higher-quality images [48]. The rates reported by other studies are showing the opposite pattern than the results in the current study, where a higher false negative and lower false positive rate is seen. This pattern is in agreement with a study looking at feature-based facial comparison [18], where fewer false positives and more false negatives were identified across photographic and CCTV images. This likely indicates the benefit of using a feature list to aid the facial comparison process [49,50] which was not done in the first two studies [4,19]. Decreasing the false positive rates across any forensic practice method is crucial to prevent wrongful convictions [51] this was achieved by including a feature list and using a stricter iteration [18] of the level of agreement system [13].

The strict iteration surpassed the lenient iteration in the digital sample in most aspects, with the exception of negative predictive value (NPV) and sensitivity which in turn marginally decreased balanced accuracy. In the analogue sample, however, even though the false positive rate decreased, most other performance measures were negatively affected in the strict iteration. Despite this shift, hit rate remained high, as did NPV and specificity. This high specificity and NPV in the analogue sample might not necessarily represent the ability of MA to correctly exclude matches in a low-quality sample as it likely results from the composition of the sample overall. This is particularly important as these values can be misconstrued as positive and misinterpreted as of value despite the overall inadequate performance of MA in low-quality image data, as shown by the extremely low chance corrected accuracy. The drastic

decrease in detection prevalence (and detection rate) throughout the analogue sample, may be indicative of the decrease in confidence to assign matches that could result in a correct decision.

The exceptional performance identified in the digital sample, is likely linked to the high amount of information available in the video recordings as well as the close correspondence of the conditions image capture, to those of the standard photograph capture. Better correspondence in view, and overall high-quality images, with minimal to no loss or distortion of information resulted in an excellent performance of MA in the digital CCTV sample. This suggests that images with the face being composed of 800 x 1000 pixels is close to ideal for facial comparison by MA with the aid of the FISWG feature list. One main concern attributable to the recordings of the digital video, is the small objective to face distance (800 mm), which is confirmed to cause perspective distortion of the edges of a 3D object and to be problematic when one-to-one comparisons of craniofacial anatomy is conducted in craniofacial superimposition [52]. Although facial comparison is different to craniofacial superimposition, the accurate presentation of facial features remains important and is still an important consideration, particularly at shorter distances [53]. At smaller objective to face distances (< 1000 mm), such as with the distance of digital video capture, the distortion experienced by the images is notable [52]; however, the standardised photographic sample was captured at 1500 mm, decreasing the distortion in the faces. Despite the high level of distortion and considering the qualitative nature of MA-based facial comparison with the lack of direct metric comparisons, this perspective distortion appeared to have limited to no effect in this sample. Furthermore, despite the occasional data loss experienced during recording due to intermittent network connectivity, no compression or any other artifacts were noted in the digital sample.

When considering the subpar performance identified in the analogue CCTV sample, the most likely attributable reason is the poor overall quality of the images extracted from the recordings. The faces shown in these images were particularly small (18 x 26 pixels) and well below the level of detail expected to be able to perform detailed facial comparisons [22]. The detail loss resulting in the face being composed of such few pixels can occur due to the limited resolution of the camera recording at a great distance [20]. The analogue camera used in the current study, recorded the individuals at approximately 2900 mm, which on the native resolution of 976 x 494 resulted in the very small representation of the face. Pixelation of the face to this degree due to low resolution recordings has been investigated in a single previous study only [20]. This study identified a sharp accuracy decrease in samples where images were pixelated, decreasing between 60% and 66%, to as low as 48% in their lowest resolution sample, while

their high-resolution sample had an 85 to 90% accuracy [20]. Interestingly, Bindemann *et al.* [20] also investigated the effect of decreasing the size of the pixelated images during the facial comparison process and found that a 75% reduction of the image size resulted in the optimum viewing conditions to attenuate the effect of pixelation and increase accuracy by 12 to 13%. In the present study, the images were viewed at their largest possible size to match the size of the standardised photographs as is common practice for forensic investigations [13], which may have further affected the accuracy of the sample in this study. This attenuating effect of decreasing the size of low-resolution images could likely have implications not only for the process of facial comparison, but also in consideration of court chart presentations. When a facial analysis expert presents their testimony in the South African legal system, they display a court chart similar in organisation to Figures 1 and 2 to explain their findings to the court [13]. A chart including low-resolution, pixelated images presented with enlarged images on a large screen may have a damaging effect on their testimony as the court may have greater difficulty than an expert in seeing the similarity or dissimilarities highlighted.

Positioning of the camera at an angle of incidence greater than 15° can limit the visibility of facial details [54]. However, in the present study's digital sample no angle of incidence was present as it was approximately at the same height as the participants faces. This positioning in conjunction with the higher quality images, may have contributed the better and more reliable performance of MA in the digital sample (97.3%) than in the suboptimal CCTV sample of chapter 3 (82.6%), which had an angle of incidence of 27°. In contrast, the analogue sample had an angle of incidence of 22°, which is greater than recommended. However, it was difficult to qualitatively determine the extent of the effect that this may have incurred in the analysis process as it is unclear from the image comparisons how much of the features that were obscured, may have been due to positioning of the camera or simply from the poor resolution (Figure 2). These findings reiterate the importance of using both high-quality and good condition recordings when conducting facial comparisons.

An additional factor intentionally introduced in the present study was monochrome images. The analogue sample severely underperformed, likely due to both low-quality and suboptimal conditions, however, it also included monochrome images. Colour is an intrinsic feature of vision and is considered to play a role in complex scene interpretation [55]. Even though many studies that investigate the effectiveness of facial comparison on CCTV recordings appear to make use of monochrome data [e.g. 17,18], none discuss lack of colour as a confounding feature in the analysis process. Psychology studies have found that presence of colour in images

can aid in detecting a face [56] and contributes to facilitating human face recognition [55], even when an image is degraded [55,56]. In the current study, the many confounding factors decreasing the performance of the analogue sample, make it difficult to isolate colour as a factor. However, when considering FISWG facial feature list, the first facial component is colour. As such, the inability to compare colour information is likely contributing to the decrease accuracy of MA's performance in the analogue sample. Interestingly, in the digital sample, where colour information was available, it often had to be disregarded as the lighting differences between the recordings and the database photographs consistently made the skin hue and tone appear different. Further studies comparing the contribution of colour in facial identification are required to truly gauge the extent of information lost in monochrome images.

1.5 Conclusion

This has been the first study to demonstrate the effects of different CCTV recordings of high and low-quality when using the FISWG feature list for facial comparison. In a high-quality and optimal IP CCTV sample, exceptional accuracies were identified with a high level of reliability and minimal false positive rates and low false negative rates, indicating outstanding performance in good conditions. Under poor-quality and suboptimal analogue CCTV conditions, morphological analysis failed to confidently match and exclude individuals correctly. This was attributed primarily to the low-resolution footage as other studies comparing face matching also found subpar performances under particularly low-quality images [19,20], as expected. Based on these observations, the authors strongly advocate against the use of MA in particularly poor-quality CCTV data, as it overall dramatically underperformed and despite a relatively low false positive rate, it retained a very high error rate with poor reliability. It is also recommended that under circumstances where the only recordings available are comparable to the low-quality analogue footage of this study, the still images extracted should be resized to a dimension that may attenuate the effect of the poor quality. However, identifying this size threshold would require further testing with footage from different quality cameras and at different image sizes.

Despite a sub-standard performance and high error rates seen in the analogue sample, hit rates appeared deceptively good, likely due to the sample distribution including a much larger number of true negative matches possible. As such, these results highlight the importance of using chance corrected accuracies and not relying on absolute hit rates as they can appear

deceptively better. Thorough consideration of sample composition used in such studies should also be undertaken to reduce the appearance of falsely promising results.

Further studies are required to identify a threshold of quality that can be considered adequate for forensic facial comparison, to confirm accurate and reliable results. In addition, detailed research investigating the various conditions of CCTV recording capture, including the angle of incidence, absence of colour, and lighting conditions is required to qualify the actual effect each of these factors have on facial comparison.

Tables

Table 3. Face pool trials by analysis: cohort subsample breakdown. Test trials were single normal trials including one target and nine foils and one correct match, No-match trials included one target opposite ten foil images. Repeat trials were only used for intra-observer analysis, Analysis trials include all trials that were statistically analysed for the majority of the study excluding the repeat trials.

Analysis Cohort	Test Trials	No-match Trials	Repeat Trials	Analysis Trials	Total trials
Digital CCTV to ST photographs	81	6	8	87	95
Analogue CCTV to ST photographs	114	8	8	122	130

ST = standard, CCTV = Closed Circuit Tele-Vision.

Table 4. Facial pose distribution throughout face pool sample.

Analysis Cohort	Anterior View	Left 45° View	Left Lateral View	Right 45° View	Right Lateral View	Total Trials
Digital CCTV to ST photographs	14	21	16	19	17	87
Analogue CCTV to ST photographs	23	25	24	26	24	122

ST = standard, CCTV = Closed Circuit Tele-Vision

Table 3: Contingency table reference for confusion matrix of positive and negative matches (adapted from Kuhn [42]). Formulae employed on each subset of analysis to compare predictive quality of analyses rely on this contingency table.

	True face match	*True face exclusion*
Matched by analyst	True Positive (TP)	False Positive (FP)
Excluded by analyst	False Negative (FN)	True Negative (TN)

Table 4: Equation list for statistical analysis based on the Caret R package [41,42].

Statistic	Equation
Sensitivity (Recall)	$\frac{TP}{(TP + FN)}$
Specificity	$\frac{TN}{(FP + TN)}$
Prevalence	$\frac{(TP + FN)}{(TP + FP + FN + TN)}$
Positive Predictive Value (Precision)	$\frac{(Sensitivity \times Prevalence)}{\{(Sensitivity \times Prevalence) + [(1 - Specificity) \times (1 - Prevalence)]\}}$
Negative Predictive Value	$\frac{[Specificity \times (1 - Prevalence)]}{\{[1 - Sensitivity) \times Prevalence] + [Specificity \times (1 - Prevalence)]\}}$
Detection Rate	$\frac{TP}{(TP + FP + FN + TN)}$
Detection Prevalence	$\frac{(TP + FP)}{(TP + FP + FN + TN)}$
Balanced Accuracy	$\frac{(Sensitivity + Specificity)}{2}$
F_1-Score	$\frac{(1 + \beta^2) \times Precision \times Sensitivity}{(\beta^2 \times Precision) + Sensitivity}$

TP = true positive; FP = false positive; TN = true negative; FN = false negative

Table 5: Tabulation of intra- and inter-observer agreements across cohorts of analysis.

	Intra-observer agreement	**p-value**	**Inter-observer agreement**	**p-value**
Digital CCTV to ST photographs	0.969	< 0.001	0.773	< 0.001

Analogue CCTV to ST photographs	0.529	< 0.001	0.378	< 0.001

Table 6: Contingency table reference for confusion matrix of the lenient iteration of the Digital CCTV cohort. Analysed sample is contrasted to the ideal perfect matches and exclusions for direct comparison of performance of the facial comparisons.

	Actual face match	*Actual face exclusion*		*Actual face match*	*Actual face exclusion*
Matched by analyst	TP = 81 (TPR= 100%)	FP = 9 (FPR= 1.1%)	*Ideal perfect match*	TP = 81	FP = 0
Excluded by analyst	FN = 0 (FNR = 0%)	TN = 780 (TNR= 98.9%)	*Ideal perfect exclusions*	FN = 0	TN = 789

TP = true positive; FP = false positive; TN = true negative; FN = false negative; TPR = true positive rate; FPR = false positive rate; FNR = false negative rate; TNR = true negative rate.

Table 7: Contingency table reference for confusion matrix of the lenient iteration of the Analogue CCTV cohort analysed. Analysed sample is contrasted to the ideal perfect matches and exclusions for direct comparison of performance of the facial comparisons.

	Actual face match	*Actual face exclusion*		*Actual face match*	*Actual face exclusion*
Matched by analyst	TP = 59 (TPR= 56.2%)	FP = 47 (FPR= 4.6%)	*Ideal perfect match*	TP = 105	FP = 0
Excluded by analyst	FN = 46 (FNR= 43.8%)	TN = 979 (TNR= 95.4%)	*Ideal perfect exclusions*	FN = 0	TN = 1026

TP = true positive; FP = false positive; TN = true negative; FN = false negative; TPR = true positive rate; FPR = false positive rate; FNR = false negative rate; TNR = true negative rate.

Table 8: Confusion matrix function results and chance corrected accuracies across cohorts of analysis.

	Digital CCTV	**Analogue CCTV**
Hit rate accuracy	0.990	0.918
95% confidence interval	0.981 - 0.995	0.900 - 0.933
No information rate (NIR)	0.907	0.907
P-Value (Acc>NIR)	< 0.001	0.118
McNemar's Test P-Value	0.008	1.000
Chance Corrected Accuracy (CCA)	94.2%	51.4%

P-value (for CCA)	< 0.001	< 0.001

Table 9: Tabulation of all predictive quality values calculated across cohorts of analyses.

	Digital CCTV	Analogue CCTV
Sensitivity (Recall)	1.000	0.562
Specificity	0.989	0.954
Prevalence	0.093	0.093
Positive predictive value (PPV)	0.900	0.557
Negative predictive value (NPV)	1.000	0.955
Detection Rate (DR)	0.093	0.052
Detection Prevalence (DP)	0.104	0.094
Balanced Accuracy	0.994	0.758
F_1-Score	0.947	0.559

Table 10: Contingency table reference for confusion matrix of the strict iteration of the Digital CCTV cohort. Analysed sample is contrasted to the ideal perfect matches and exclusions for direct comparison of performance of the facial comparisons.

	Actual face match	*Actual face exclusion*		*Actual face match*	*Actual face exclusion*
Matched by analyst	TP = 79 (TPR = 97.5%)	FP = 2 (FPR = 0.3%)	*Ideal perfect match*	TP = 81	FP = 0
Excluded by analyst	FN = 2 (FNR = 2.5%)	TN = 787 (TNR = 99.7%)	*Ideal perfect exclusions*	FN = 0	TN = 789

TP = true positive; FP = false positive; TN = true negative; FN = false negative; TPR = true positive rate; FPR = false positive rate; FNR = false negative rate; TNR = true negative rate.

Table 11: Contingency table reference for confusion matrix of the strict iteration of the Analogue CCTV cohort analysed. Analysed sample is contrasted to the ideal perfect matches and exclusions for direct comparison of performance of the facial comparisons.

	Actual face match	*Actual face exclusion*		*Actual face match*	*Actual face exclusion*
Matched by analyst	TP = 26 (TPR = 24.8%)	FP = 12 (FPR = 1.2%)	*Ideal perfect match*	TP = 105	FP = 0

Excluded by analyst	FN = 79 (FNR = 75.2%)	TN = 1014 (TNR = 98.8%)	*Ideal perfect exclusions*	FN = 0	TN = 1026

TP = true positive; FP = false positive; TN = true negative; FN = false negative; TPR = true positive rate; FPR = false positive rate; FNR = false negative rate; TNR = true negative rate.

Table 12: Confusion matrix function results across the strict iteration of each cohort of analysis.

	Digital CCTV	Analogue CCTV
Hit rate accuracy	0.995	0.920
95% confidence interval	0.988 – 0.999	0.902 - 0.935
No information rate (NIR)	0.907	0.907
P-Value (Acc>NIR)	< 0.001	0.081
McNemar's Test P-Value	1.000	< 0.001
Chance Corrected Accuracy (CCA)	97.3%	33.1%
P-value (for CCA)	< 0.001	< 0.001

Table 13: Tabulation of the predictive quality values calculated across the strict iteration of each cohort of analysis.

	Digital CCTV	Analogue CCTV
Sensitivity (Recall)	0.975	0.248
Specificity	0.998	0.988
Prevalence	0.093	0.093
Positive predictive value (PPV)	0.975	0.684
Negative predictive value (NPV)	0.998	0.928
Detection Rate (DR)	0.091	0.023
Detection Prevalence (DP)	0.093	0.034
Balanced Accuracy	0.986	0.618
F_1-Score	0.975	0.364

Figures

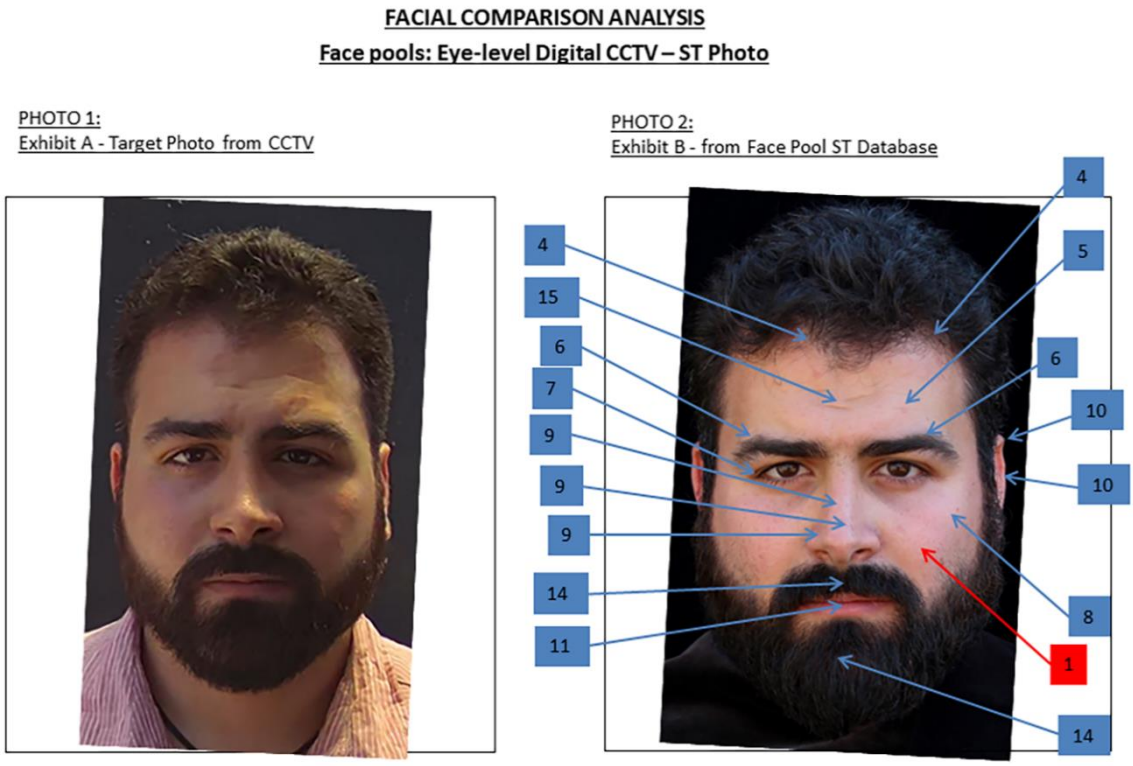


Figure 1: Facial comparison analysis example from the Digital CCTV sample. A target image from CCTV was placed next to each possible match based on possible positive matching during holistic analysis. Standardised (ST) photograph (right) was compared to the target image from the Digital CCTV recordings (left), using the FISWG facial component list [35]. Similar (blue) and dissimilar (red) facial components were highlighted by specific facial component number from the FISWG list.

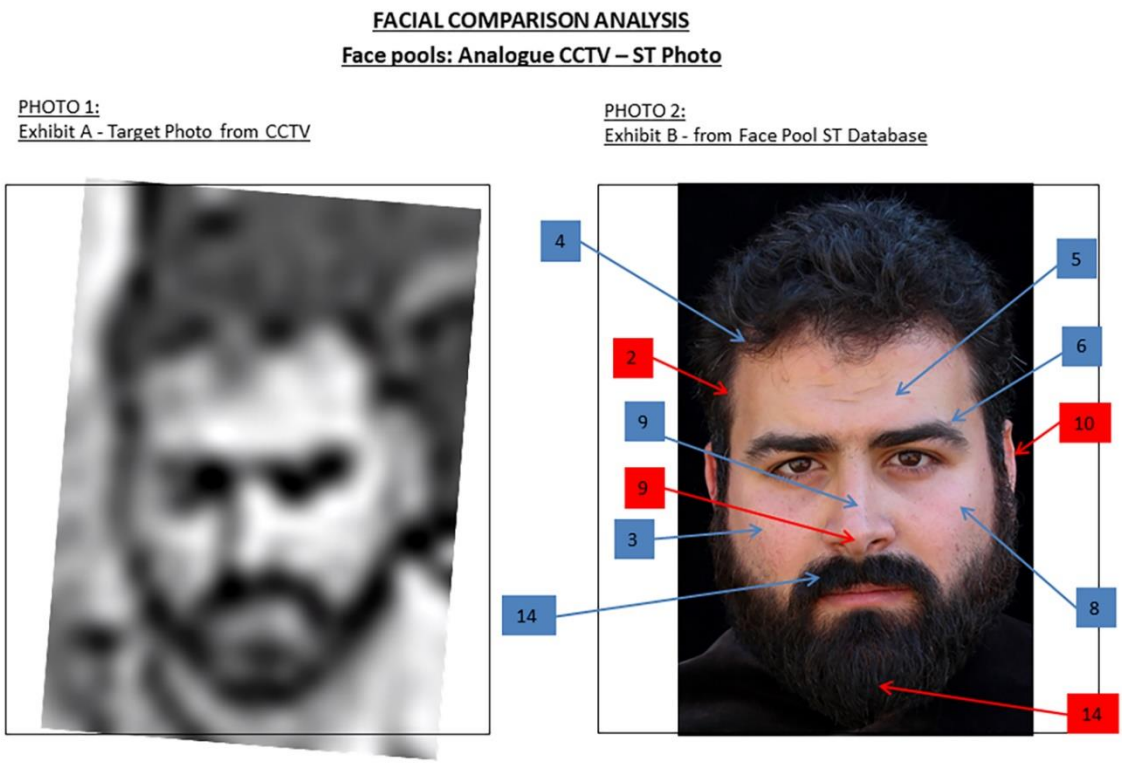


Figure 2: Facial comparison analysis example from the analogue CCTV sample. A target image from CCTV or wildtype (WT) photographs was placed next to each possible match based on possible positive matching during holistic analysis. Standardised (ST) photograph (right) was compared to the target image from the analogue CCTV recordings (left), using the FISWG facial component list [35]. Similar (blue) and dissimilar (red) facial components were highlighted by specific facial component number from the FISWG list.



Figure 3: Bar graphs of the facial comparison outcomes under a lenient (A) and strict (B) iteration within the Digital CCTV to photograph cohort by trial. TP = true positive; FP = false positive; TN = true negative; FN = false negative; IA = inconclusive analysis.

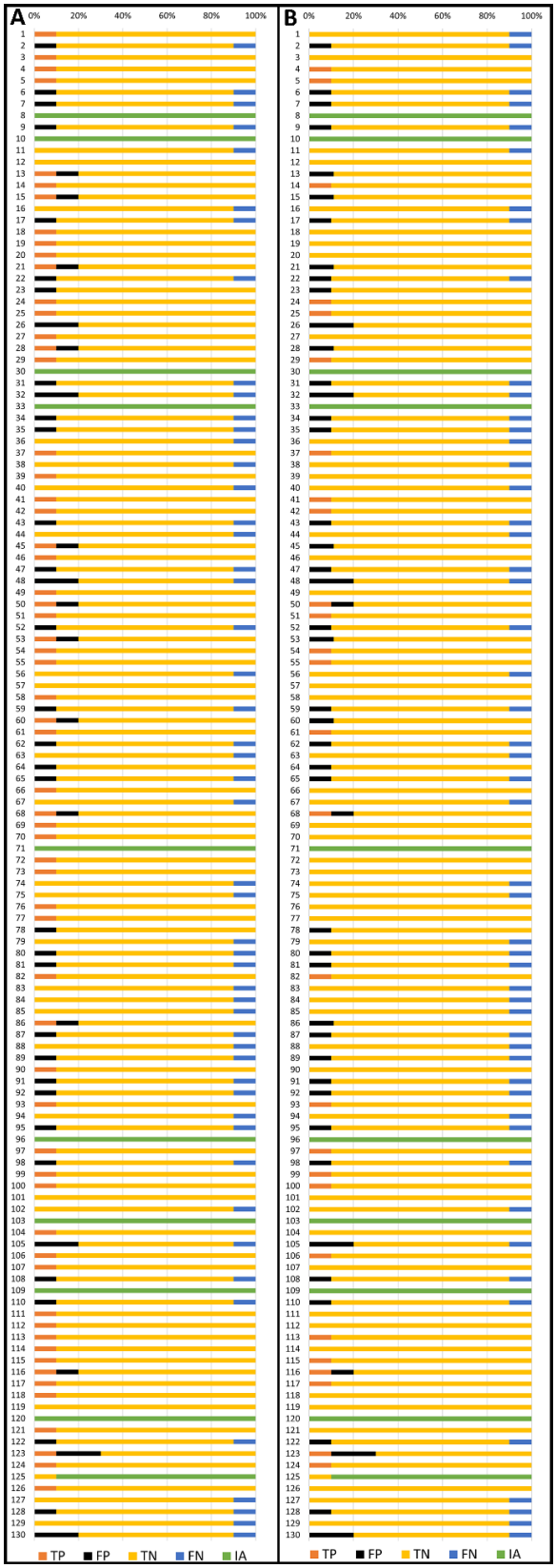


Figure 4: Bar graphs of the facial comparison outcomes under a lenient (A) and strict (B) iteration within the Analogue CCTV to photograph cohort by trial. TP = true positive; FP = false positive; TN = true negative; FN = false negative; IA = inconclusive analysis.

References

[1] T. Valentine, J.P. Davis, Forensic facial identification: theory and practice of identification from eyewitnesses, composites and CCTV, John Wiley & Sons, Ltd, Chichester, UK, 2015. https://doi.org/10.1002/9781118469538.

[2] A.K. Jain, B. Klare, U. Park, Face matching and retrieval in forensics applications, IEEE Multimed. 19 (2012) 20–27. https://doi.org/10.1109/MMUL.2012.4.

[3] H. Walker, A. Tough, Facial Comparison from CCTV footage: The competence and confidence of the jury, Sci. Justice. 55 (2015) 487–498. https://doi.org/10.1016/j.scijus.2015.04.010.

[4] C. Wilkinson, R. Evans, Are facial image analysis experts any better than the general public at identifying individuals from CCTV images?, Sci. Justice. 49 (2009) 191–196. https://doi.org/10.1016/j.scijus.2008.10.011.

[5] C. Norris, M. McCahill, D. Wood, Editorial. The growth of CCTV: A global perspective on the international diffusion of video surveillance in publicly accessible space, Surveill. Soc. 2 (2004) 110–135.

[6] D.J. Robertson, M.C. Fysh, M. Bindemann, Facial identity verification: Five challenges facing practitioners, Keesing J. Doc. (2019) 1–8.

[7] V. Bruce, Z. Henderson, K. Greenwood, P.J.B. Hancock, A.M. Burton, P. Miller, Verification of face identities from images captured on video, J. Exp. Psychol. Appl. 5 (1999) 339–360. https://doi.org/10.1037/1076-898X.5.4.339.

[8] M. Burton, D. White, A. McNeill, The Glasgow face matching test, Behav. Res. Methods. 42 (2010) 286–291. https://doi.org/10.3758/BRM.42.1.286.

[9] F.A. Pavel, E. Iordănescu, The influence of facial expressions on recognition performance in facial identity, Procedia - Soc. Behav. Sci. 33 (2012) 548–552. https://doi.org/10.1016/j.sbspro.2012.01.181.

[10] R. Jenkins, A.M. Burton, Limitations in facial identificiation: The evidence, Justice of the Peace. 172 (2008) 4–6.

[11] J.H. Searcy, J.C. Bartlett, A. Memon, Age differences in accuracy and choosing in eyewitness identification and face recognition, Mem. Cogn. 27 (1999) 538–552.

https://doi.org/10.3758/BF03211547.

[12] M.H. Papesh, Photo ID verification remains challenging despite years of practice, Cogn. Res. Princ. Implic. (2018). https://doi.org/10.1186/s41235-018-0110-y.

[13] M. Steyn, M. Pretorius, N. Briers, N. Bacci, A. Johnson, T.M.R. Houlton, Forensic facial comparison in South Africa: State of the science, Forensic Sci. Int. 287 (2018) 190–194. https://doi.org/10.1016/J.FORSCIINT.2018.04.006.

[14] Facial Identification Scientific Working Group, Facial Comparison Overview and Methodology Guidelines, 2019. https://fiswg.org/fiswg_facial_comparison_overview_and_methodology_guidelines_V1.0_20191025.pdf.

[15] A.M. Megreya, M. Bindemann, C. Havard, Sex differences in unfamiliar face identification: Evidence from matching tasks, Acta Psychol. (Amst). 137 (2011) 83–89. https://doi.org/10.1016/j.actpsy.2011.03.003.

[16] G. Porter, G. Doran, An anatomical and photographic technique for forensic facial identification, Forensic Sci. Int. 114 (2000) 97–105. https://doi.org/10.1016/S0379-0738(00)00290-5.

[17] V. Bruce, Z. Henderson, C. Newman, A.M. Burton, Matching identities of familiar and unfamiliar faces caught on CCTV images, J. Exp. Psychol. Appl. 7 (2001) 207–218. https://doi.org/10.1037//1076-898X.7.3.207.

[18] N. Bacci, T.M.R. Houlton, N. Briers, M. Steyn, Validation of forensic facial comparison by morphological analysis in photographic and CCTV samples, Int. J. Legal Med. (2021). (In Press)

[19] W.-L. Lee, C. Wilkinson, A. Memon, K. Houston, Matching unfamiliar faces from poor quality closed-circuit television (CCTV) footage: An evaluation of the effect of training on facial identification ability, AXIS Online J. Cent. Anat. Hum. Identif. 1 (2009) 19–28.

[20] M. Bindemann, J. Attard, A. Leach, R.A. Johnston, The effect of image pixelation on unfamiliar-face matching, Appl. Cogn. Psychol. 27 (2013) 707–717. https://doi.org/10.1002/acp.2970.

[21] D. Gibelli, Z. Obertová, S. Ritz-Timme, P. Gabriel, T. Arent, M. Ratnayake, D. De Angelis, C. Cattaneo, The identification of living persons on images: A literature review,

Leg. Med. 19 (2016) 52–60. https://doi.org/10.1016/j.legalmed.2016.02.001.

[22] V. Damjanovski, CCTV from light to pixels, 3rd ed., Elsevier, Oxford, 2014.

[23] A.M. Burton, S. Wilson, M. Cowan, V. Bruce, Face recognition in poor-quality video: evidence from security surveillance, Psychol. Sci. 10 (1999) 243–248. https://doi.org/10.1111/1467-9280.00144.

[24] R. Surette, The thinking eye, Polic. An Int. J. Police Strateg. Manag. 28 (2005) 152–173. https://doi.org/10.1108/13639510510581039.

[25] S. Smith, CCTV market outlook 2017, Cision PR Newswire. (2014). https://www.prnewswire.com/news-releases/cctv-market-outlook-2017-259397561.html.

[26] 6Wresearch, South Africa video surveillance market (2016-2022), 2016. https://www.6wresearch.com/industry-report/south-africa-video-surveillance-market-by-revenue-type-share-size-forecasts-growth-opportunity-trends-players.

[27] L. Wood, CCTV Cameras - Worldwide Market Outlook Report 2018-2026 : Dome Cameras Dominate, Businesswire. (2018). https://www.businesswire.com/news/home/20180913005519/en/CCTV-Cameras---Worldwide-Market-Outlook-Report (accessed September 23, 2020).

[28] Research and Markets, Global CCTV Market Forecast 2022, (2018). https://www.researchandmarkets.com/reports/3512790/global-cctv-market-forecast-2022#:~:text=The CCTV market is gaining,concerns for security and safety.&text=In their latest research study,11%25 during 2018-2022. (accessed September 23, 2020).

[29] M.J. Oosthuizen, What to look out for in 2014, Hi-Tech Secur. Solut. (2014). http://www.securitysa.com/6411R.

[30] N.N.A.N. Ghazali, N.A. Zamani, S.N.H.S. Abdullah, J. Jameson, Super resolution combiantion methods for CCTV forensic interpretation, in: 12th Int. Conf. Intell. Syst. Des. Appl., 2012: pp. 853–858.

[31] S. Viték, M. Klíma, L. Krasula, Video compression technique impact on efficiency of person identification in CCTV systems, in: Proc. - Int. Carnahan Conf. Secur. Technol., Rome, Italy, 2014. https://doi.org/10.1109/CCST.2014.6987027.

[32] X. Qi, C. Liu, Mitigate compression artifacts for face in video recognition, in: M.

Blowers, R.D. Hall, V.R. Dasari (Eds.), Disruptive Technol. Inf. Sci. IV, SPIE, 2020: p. 25. https://doi.org/10.1117/12.2556702.

[33] M.C. Fysh, M. Bindemann, The Kent Face Matching Test, Br. J. Psychol. 109 (2018) 219–231. https://doi.org/10.1111/bjop.12260.

[34] H.U. Keval, M.A. Sasse, Can we ID from CCTV: Image quality in digital CCTV and face identification performance, in: Def. Secur. Symp., 2008: pp. 69820K-1–15. https://doi.org/10.1117/12.774212.

[35] Facial Identification Scientific Working Group, Facial Image Comparison Feature List for Morphological Analysis, 2018. https://fiswg.org/FISWG_Morph_Analysis_Feature_List_v2.0_20180911.pdf.

[36] [dataset] N. Bacci, J. Davimes, M. Steyn, N. Briers, Wits Face Database, Wits Institutional Repos. Environ. Dsp. (2020). https://doi.org/http://dx.doi.org.10.18772/10539/20690.

[37] N. Bacci, J. Davimes, M. Steyn, N. Briers, Development of the Wits Face Database: an African database of high-resolution facial photographs and multimodal closed-circuit television (CCTV) recordings, F1000Research. (2021) (Submitted)

[38] Facial Identification Scientific Working Group, FISWG Guidelines for Facial Comparison Methods, 2012. papers3://publication/uuid/D3938836-66E9-4E88-8A1D-A0143D106FCE.

[39] R Core Team, R: a language and environment for statistical computing, (2020). https://www.r-project.org/.

[40] P.F. Watson, A. Petrie, Method agreement analysis: A review of correct methodology, Theriogenology. 73 (2010) 1167–1179. https://doi.org/10.1016/j.theriogenology.2010.01.003.

[41] M. Kuhn, Caret: classification and regression training R package, (2020). https://cran.r-project.org/package=caret.

[42] M. Kuhn, Building predictive models in R using the caret package, J. Stat. Softw. 28 (2008) 1–26. https://doi.org/10.18637/jss.v028.i05.

[43] T.G. Dietterich, Approximate statistical tests for comparing supervised classification learning algorithms, Neural Comput. 10 (1998) 1895–1923. https://doi.org/10.1007/978-3-

319-50926-6_6.

[44] FISWG, Facial Image Comparison Feature List for Morphological Analysis, (2013) 1–27.

[45] Z. Henderson, V. Bruce, A.M. Burton, Matching the faces of robbers captured on video, Appl. Cogn. Psychol. 15 (2001) 445–464. https://doi.org/10.1002/acp.718.

[46] S. Ritz-Timme, P. Gabriel, Z. Obertovà, M. Boguslawski, F. Mayer, A. Drabik, P. Poppa, D. De Angelis, R. Ciaffi, B. Zanotti, D. Gibelli, C. Cattaneo, A new atlas for the evaluation of facial features: advantages, limits, and applicability, Int. J. Legal Med. 125 (2011) 301–306. https://doi.org/10.1007/s00414-010-0446-4.

[47] M.M. Roelofse, M. Steyn, P.J. Becker, Photo identification: Facial metrical and morphological features in South African males, Forensic Sci. Int. 177 (2008) 168–175. https://doi.org/10.1016/j.forsciint.2007.12.003.

[48] J.P. Davis, T. Valentine, C. Wilkinson, Facial image comparison, in: C. Wilkinson, C. Rynn (Eds.), Craniofacial Identif., 1st ed., Cambridge University Press, New York, 2012: pp. 136–153. https://doi.org/10.1017/CBO9781139049566.015.

[49] A. Towler, D. White, R.I. Kemp, Evaluating the feature comparison strategy for forensic face identification, J. Exp. Psychol. Appl. 23 (2017) 47–58. https://doi.org/10.1037/xap0000108.

[50] A.M. Megreya, M. Bindemann, Feature instructions improve face-matching accuracy, PLoS One. 13 (2018) 1–16. https://doi.org/10.1371/journal.pone.0193455.

[51] W.C. Thompson, F. Taroni, C.G.G. Aitken, How the Probability of a False Positive Affects the Value of DNA Evidence, J. Forensic Sci. 48 (2003) 2001171. https://doi.org/10.1520/jfs2001171.

[52] C.N. Stephan, Perspective distortion in craniofacial superimposition: Logarithmic decay curves mapped mathematically and by practical experiment, Forensic Sci. Int. 257 (2015) 520.e1-520.e8. https://doi.org/10.1016/j.forsciint.2015.09.009.

[53] C.N. Stephan, J.M. Caple, P. Guyomarc'h, P. Claes, An overview of the latest developments in facial imaging, Forensic Sci. Res. 4 (2019) 10–28. https://doi.org/10.1080/20961790.2018.1519892.

[54] D. Ward, Testing Camera Height vs Image Quality, Pennsylvania, USA, 2013. https://ipvm.com/reports/testing-camera-height.

[55] A.W. Yip, P. Sinha, Contribution of color to face recognition, Perception. 31 (2002) 995–1003. https://doi.org/10.1068/p3376.

[56] M. Bindemann, A.M. Burton, The role of color in human face detection, Cogn. Sci. 33 (2009) 1144–1156. https://doi.org/10.1111/j.1551-6709.2009.01035.x.

5. Assessing the effect of facial disguises on forensic facial comparison by morphological analysis

Nicholas Bacci, Nanette Briers, Maryna Steyn

Journal of Forensic Sciences, 2021 (Accepted – In Press)

https://doi.org/10.1111/1556-4029.14722

Title: Assessing the effect of facial disguises on forensic facial comparison by morphological analysis

Running Head: Effects of disguises on forensic facial comparison

Nicholas Bacci[1], Nanette Briers[1], Maryna Steyn[1]

[1]Human Variation and Identification Research Unit, School of Anatomical Sciences, University of the Witwatersrand, Johannesburg, South Africa

Corresponding author:

Mr N Bacci, MSc (ORCID: 0000-0001-6858-7598)

Human Variation and Identification Research Unit, School of Anatomical Sciences, University of the Witwatersrand, 7 York Road, Parktown 2193, South Africa

Email: nicholasbacci@gmail.com and nicholas.bacci@wits.ac.za

Prof M Steyn, PhD (ORCID: 0000-0002-0215-9723)

Dr N Briers, PhD (ORCID: 0000-0002-8544-6049)

Declarations

Funding: The current study was conducted with support from the South African National Research Foundation (NRF) and the J.J.J. Smieszek Fellowship from the School of Anatomical Sciences, University of the Witwatersrand (DAAD-NRF and J.J.J. Smieszeck Fellowship funds awarded to N. Bacci (Grant No.: 11858) and NRF funds awarded to N. Briers as part of the Improving Methodologies and Practices in Craniofacial Identification (Grant No.: CSUR160425163022; UID: 106031). Any opinions, findings and conclusions or recommendations expressed in this study are those of the authors and therefore the NRF and University of the Witwatersrand, Johannesburg, do not accept any liability in regard thereto.

Conflicts of interest: None.

Ethics approval: Approval for the present study was obtained from the Human Research Ethics Committee (Medical) of the University of the Witwatersrand, Johannesburg. Ethics clearance certificate number: M171026.

Acknowledgements

Acknowledgement is owed to all the participants who agreed to be photographed and recorded for the development of the Wits Face Database from which the face pool samples used in the current were composed. Thanks are due to Joshua Davimes for his instrumental work in the development of the Wits Face Database, Gideon LeRoux for his assistance in the capturing and extracting of the CCTV recordings from the University security systems. Thanks are owed to Tamara Lottering for her assistance in composing the face pools and Tobias Houlton for aiding in study validation. Thanks are also due to the volunteers who assisted in participant recruitment: Jesse Fredericks, Kiveshen Pillay, Rethabile Masiu, Sameerah Sallie, Daniel Munesamy, Laurette Joubert, Jordan Swiegers, Betty Mkabela, Johannes P. Meyer, Amy Spies, Natasha Loubser, Nicole Virgili, Dan-Joel Lukumbi, Tamara Lottering, Mathabatha Ntjie, Claudia Landsman, Raheema Dalika, Merete Goosen, Stephanie Souris, Rabelani Negota, Mahlatse Mahasha, Jessica Manavhela.

ABSTRACT

Disguises are commonly used to mask a person's facial appearance in areas under closed-circuit television (CCTV) surveillance. While many studies attempted to understand the effects of disguises, such as hats and glasses, on facial recognition, limited studies have looked at disguises in forensic facial comparison. The aim of this study was to compare the outcomes of forensic facial comparison by morphological analysis (MA) in a CCTV sample with sunglasses and brimmed caps. The sample was obtained from the Wits Face Database and organised into 81 face pools of one target facial image wearing a disguise (cap or sunglasses) and 10 potential matching images. MA was conducted across face pools and confusion matrices were used to assess the outcomes. Surprisingly, sunglasses had limited effect on MA performance both in accuracy (90.4%) and reliability ($\kappa = 0.798$) while caps markedly decreased both accuracy (68.1%) and reliability ($\kappa = 0.639$). Error rates were associated primarily with false negatives in both samples (caps: 42.4%; sunglasses: 16.1%) despite the sample distribution favouring false positive errors, which were very low (caps: 0.6%; sunglasses: 0%). Similarly to other studies, hats and caps were more harmful to correct identification when compared to sunglasses, which actually resulted in better accuracy than regular CCTV recordings. The

effect of brimmed caps on accuracy was attributed to the overall loss of facial information caused. On training analysts, it may be helpful to instruct purposefully avoiding overreliance on easily disguised facial features, as other regions of the face also contain substantial feature information.

KEYWORDS: forensic facial comparison, morphological analysis, CCTV, caps, glasses, disguise, facial identification

HIGHLIGHTS

- The effects of facial disguises are understudied in forensic facial comparison.
- This is the first study to test morphological analysis on disguised faces in a CCTV setting.
- Brimmed caps were found to hide more facial information than sunglasses and decrease accuracy most.
- Sunglasses were found to have minimal effects on feature-based facial comparison.

Introduction

Closed-circuit television (CCTV) surveillance is common in both public and private areas (1–3). CCTV recordings and images extracted from these recordings are often used for facial identification purposes (4), particularly in criminal proceedings. Unfamiliar face identification is widely considered a complex task (5–7), with experts performing relatively better than the general public (8). As a result, many prosecutors rely on expert assessments and testimony for facial comparison cases (9,10).

Despite the many studies investigating human cognitive ability to correctly identify individuals' facial images (e.g. 11–14), few take into consideration external confounding factors such as disguises (14–17). It has long been accepted that various forms of disguises hamper the ability to identify faces (18), despite the occasional inconsistent findings of their effects on identification (19). Disguises, in the form of sunglasses and hats, are often worn by criminals and are widely considered effective in obscuring facial appearance and maintaining anonymity (10,20). However, these effects are relatively disregarded in facial comparison research and have not been investigated extensively in forensic facial comparison.

Eyeglasses or sunglasses and hats or caps are some of the most commonly used disguises affecting facial recognition and identification (e.g. 15,21,22). The majority of studies investigating disguises, do so in relation to some form of facial memory, often in the context of eyewitness identification (e.g. 19,23,24), hence looking at faces with a level of familiarity. In a setting where CCTV evidence is analysed by a facial comparison expert, faces are predominantly unfamiliar to the analyst performing the comparison, which relies on a different cognitive approach (12,25). In fact, many studies suggest that when attempting familiar face recognition, priority is given to internal facial features, such as the eyes and nose, as opposed to peripheral features, such as face contour shape and ear morphology (25,26). Others have suggested that familiar faces are recognised based on an average of the variations in their appearance (27), with further postulation that familiarisation creates an understanding of the truly invariable features of each face (28). Nonetheless, understanding which facial features are prioritised in unfamiliar face recognition is particularly relevant when disguises are involved, as it means that a further limitation on available features for facial comparison is imposed.

When conducting facial comparison, a systematic approach is recommended, particularly with the use of a facial feature list as it has been shown to improve accuracy (29–32). Facial

comparison in the forensic context is usually conducted under the recommendations proposed by the Facial Identification Scientific Working Group (FISWG), which has established policies and guidelines for practice and research in facial identification and facial recognition (33). According to FISWG (31) and research in the field of facial identification (e.g. 8,10,34,35), the recommended method for forensic facial comparison is considered to be morphological analysis with the aid of feature lists or instructions (29,30,36). The most recent and extensive feature list developed for this purpose was put forward by FISWG in 2018 (37). This feature list includes detailed descriptors regarding each individual anatomical facial feature to provide a thorough guide during analysis. This FISWG feature list has recently been tested and was found reliable in both photographic and CCTV derived images of relatively good quality (32). Although multiple studies have looked at disguises in general face recognition (e.g. 14–17), to date, feature list based facial comparison has not been tested in the context of disguised faces and it has been recommended that further research is required in suboptimal conditions (38). Conducting thorough testing of methodologies is particularly relevant in the field of facial comparisons as its admissibility has been questioned due to lack of testing and validation (39).

The aim of this study was, hence, to test the effectiveness of facial comparison by morphological analysis using the most extensive available feature list (37) in CCTV footage with individuals wearing face obscuring accessories, namely brimmed caps and sunglasses. The study hypothesised that both caps and sunglasses would negatively impact the ability to identify faces using morphological analysis.

Materials and Methods

This study made use of standardised (ST) photographs and CCTV recordings originating from the Wits Face Database (40,41). See Bacci et al. (41) for further details on database composition, design and image data acquisition. Still images from the recordings and standardised photographs were used to compile face pools from a total of 86 consenting, known individual, young male adults of black South African descent. From this database, face pools were composed which included one still target image and 10 prospective matches/foils from the ST photographs. Of the 10 prospective matches, nine were incorrect matches in most of the face pools; however, a few face pools included 10 incorrect prospective matches so that the analyst did not feel obliged / subjectively pressurized to make a match in every analysis (Table 1). These face pools were compiled by someone other than the primary analyst to ensure

blinding during analysis. Across all face pools, the maximum number of repetitions of an image permitted was 15, in order to prevent the analyst from over familiarisation with the participants' faces.

The target images were extracted from a pre-installed CCTV Internet Protocol (IP) camera (HikVision, model: DS-2CD2142FWD-I, 4-megapixel) with an approximate objective to face distance of 3030 mm and at an approximate 27° angle of incidence. Wits Face Database CCTV recordings were captured over approximately 30 seconds with participants standing at ease, facing straight ahead, with a neutral expression across five facial views (anterior, left 45°, left lateral, right 45° and right lateral) for approximately five seconds in each view (41). A facial image was extracted from each of these five views recorded for each participant. Participants included in the subset of the database utilised for this study were recorded wearing a disguise, either in the form of sunglasses (Glasses CCTV) or a brimmed cap (Cap CCTV) (FIG. 1 and FIG. 2). No figures of database participants are allowed to be published due to ethical restrictions, as such, images included for visualisation purposes originate from the freely available sample of the Wits Face Database (https://hdl.handle.net/10539/29924) (40,41). The cap and sunglasses used were identical in all cases to prevent creating additional variation in the footage. ST photographs were obtained from a Canon 1300D 18MP DSLR camera (18-55 mm DC Canon lens) at an image resolution of 3456 x 4320 pixels at 300 dpi, while at an objective to face distance of 1500 mm (41). The participants' clothing and the background were obscured by a black cloth and no disguise items were included in the standardised photos. Both still images and ST photographs were cropped at a 4:5 aspect ratio with the face centred in each, resulting in an approximate CCTV still image resolution of 368 x 460 pixels at 96 dpi (41). The ST photographs were captured at the same five facial views indicated above for the CCTV images but at an angle of incidence of approximately 0° (41). These images were meant to represent realistic conditions of CCTV data capture and images captured with a good quality photographic camera for comparison (41). The CCTV recordings were directly extracted from a pre-existing security system installation and the photographs were obtained from a low- to mid-range camera akin to what police forces would use.

The face pools were used for test trials, analysis trials, repeat trials and no-match trials (Table1). Test trials were single normal trials including one target and nine foils and one correct match. No-match trials included one target and ten foil images. Repeat trials were only used for intra-observer analysis. Analysis trials included all trials that were statistically assessed for the majority of the study excluding the repeat trials. For inter-observer testing a total of 10

analysis trials (5 with caps and 5 with sunglasses) were analysed by a forensic facial comparison expert and compared to the principal analyst's comparisons. Repeat trials were distributed evenly throughout each cohort to prevent familiarisation during facial comparison, while being separated by at least 10 trials from one another in the entire analysis face pool group and analysed over different days. In addition to this trial distribution, a relatively even distribution of views was compiled in the face pools to allow a varied distribution of facial features (Table 2) to be compared. Although it can be argued that the selection of a specific facial view would be preferable to permit a more consistent analysis, it would also pose a bias towards specific facial features clearly identifiable in the selected facial view during analysis. It was therefore decided to use a variety of facial views.

Analysis of the face pools was conducted following the FISWG guidelines on facial comparison methods (31), as also described in Bacci *et al.* (32). No part of this process was automated as the morphological analysis process is entirely human observer-based. This process included an initial holistic review of the target image contrasted to each possible match in a face pool to eliminate clear mismatches, followed by more detailed morphological analysis with all potential matches identified in the holistic review phase. This analysis process was not carefully timed; however, each face pool analysis procedure required between one to two hours to analyse. Particularly obvious mismatches were usually discarded in less than five seconds and any potential matches fully analysed in the timeframe of one to two hours, depending on the number of possible matches identified within a specific pool. Most face pools included at least one possible match (n = 41), a few included two possible matches (n = 27), and only a handful included three or more matches (n = 9) and fewer included no possible matches at all (n = 8). The outcomes included a prospective possible match, not a match or no analysis possible. All comparisons considered prospective possible matches were evaluated in detail using morphological analysis following the FISWG facial comparison feature list for morphological analysis summarised in Table 3. Morphological analysis follows an analysis, comparison, evaluation, and verification (ACE-V) methodology (31), which is traditionally employed in various forensic practices to ensure a systematic, staged process is followed (42). Facial comparisons by morphological analysis were conducted by comparing the target and ST images adjacent to one another in Microsoft PowerPoint (365, v. 16), following their adjustment to the same rotation and size without altering the aspect ratio or other intrinsic properties of the images (e.g. brightness, contrast etc.) (FIG. 1 and FIG. 2). See Bacci *et al.* (32) for more detail. Facial features were then compared individually, and corresponding

features were highlighted using their numerical allocation from the FISWG feature list. A final decision was then made for each of the facial comparisons by using an adapted version of the classifications and definitions employed by the South African Police Service (43). Once final decisions were made based on these classifications, the scores representing a strong positive conclusion (1) and an inconclusive conclusion with a level of morphological similarity (2) were considered a match, while inconclusive with some holistic similarity (3) and negative (4) conclusions were considered not a match. Any analyses that could not be performed with any level of confidence were considered inconclusive (5) and excluded from further analysis. These scores were then compared to the true match-up of the face pools to calculate the overall hit rate for each cohort of analysis. These outcomes were categorised as true positive (TP), a false positive (FP), a false negative (FN), a true negative (TN), or an inconclusive analysis (IA) based on their correspondence with the true match-up.

Statistical analyses were conducted in R Studio (version 3.6.3) (44) using the Caret and the irr R packages (45–47). Initial match-up hit rate calculation was done during data management using Microsoft Excel (365, v. 16) and chance mitigation was performed via Cohen's kappa to determine chance corrected accuracy of morphological analysis in each cohort. A kappa value greater than 0.80 (48) was considered agreeable for a specific cohort. Intra-observer agreement was also tested using weighted Cohen's kappa within each match-up decision of the repeat trials. Based on the individual comparison hit rates, a contingency table was compiled for each cohort of analysis (Table 4).

Based on this contingency table, the confusion matrix function from the Caret package was used to calculate a series of measures of the predictive power of the analysis (Table 5). A raw overall accuracy is calculated by dividing the total correct scores over the total sample, with a 95% confidence interval. The raw accuracy was then compared to the No Information Rate (NIR), defined as the largest fraction of the observed outcomes (46) from which a p-value was calculated, to verify whether the raw hit rate accuracy was statistically above than the NIR (47). The ratio of error types (false positives vs false negatives) within a specific cohort was determined by conducting a McNemar's test, where p-value lower than $\alpha = 0.05$ indicated a difference in this proportion of errors (49), with errors skewing towards false positive or false negative types.

The equations used to calculate the predictive efficacy of morphological analysis within each cohort were from the Caret R package (46,47) and are described in Table 5 and the full equations are available as supplementary information. These analyses were conducted at two

different thresholds to compare the performance of morphological analysis within a lenient and a strict iteration. Within the lenient iteration, comparisons scored as both positive matches and possible matches based on holistic resemblance (scores of 1 and 2) were assigned as matches, while inconclusive analyses with some holistic similarity and no match (scores of 3 and 4) were assigned as not matching. In the strict iteration, only a clear positive match (score of 1) was considered as matching, while the more inconclusive or possible matches were grouped with the "no match" group (scores of 2, 3 and 4). This variation in threshold of acceptance of a match was intended to allow for only the highest level of confidence-based matches to be considered as truly positive, in an attempt at reducing false positive errors.

Results

Intra-observer reliability fluctuated between the two sample (caps and sunglasses) cohorts, with an almost perfect agreement in the sunglasses CCTV cohort ($\kappa = 0.798$; $p < 0.001$) and a lower yet still substantial agreement in the cap CCTV cohort ($\kappa = 0.639$; $p < 0.001$). Inter-observer agreement was found to be substantial in the sunglasses group of analysis ($\kappa = 0.702$; $p < 0.001$) and moderate for the brimmed caps group ($\kappa = 0.557$; $p < 0.001$). The individual outcomes for each face pool, with the exclusion of the repeat face pools used for intra-observer agreement testing, are shown for both Cap CCTV (FIG. 3) and Sunglasses CCTV (FIG. 4). From the sunglasses cohort, all facial comparisons were included in the analysis as none were found to be not scorable, while in the cap cohort a total of eight comparisons from one trial were classified as inconclusive / unscorable and excluded. In the target image of these comparisons the cap obscured enough facial detail to make facial comparison impossible.

The confusion matrices indicated a notably stronger performance of morphological analysis in the sunglasses CCTV cohort as opposed to the cap cohort (Tables 6 and 7). Hit rate accuracy (HA) was similarly high for both the cap (0.917, CI = 0.884 - 0.943) and sunglasses (0.979, CI = 0.958 - 0.992) cohorts; however, this accuracy was not significantly higher than the NIR for the cap cohort (Table 8). This result suggests that morphological analysis might not be particularly effective in CCTV recordings with brimmed caps. With regard to proportion of errors, cap CCTV appeared to have a similar distribution of error rates (false positive and false negative errors) (McNemar p-value = 0.018), while in sunglasses errors were more skewed towards false positives in raw numbers, despite the overall lower false positive rate (FPR) (1.9%) (Table 6). The FPR was lower in comparison to the false negative rate (FNR) in the

sunglasses cohort (3.2%) and the FPR for the cap cohort (6.7%), which is indicative of an overall better quality of matching in the sunglasses sample. Due to sample composition there is a greater potential for false positive outcomes, which is seen in the raw higher number of FP over FN outcomes (Tables 6 and 7). Despite this the FNR was consistently higher than the FPR across both samples.

Balanced accuracies (BA) for both cohorts were high, with the sunglasses sample (97.4%) affecting MA less than the cap sample (84.5%) (Table 8), suggesting a slightly better accuracy in the sunglasses CCTV when considering sensitivity and specificity of the two samples. However, the chance corrected accuracy (CCA) for the sunglasses cohort (88.4%) was notably higher than the cap cohort (58.0%) and above the acceptable threshold of 0.80 (48). This, in turn, highlights the concerningly stronger effect of caps on MA accuracy. This reduced accuracy is further highlighted by the markedly lower sensitivity within the cap sample (0.758) over the sunglasses sample (0.968), despite the identical prevalence score (0.091) (Table 9). As a result, the positive predictive value (PPV) of MA in the cap cohort was also severely reduced (0.532) compared to the sunglasses sample (0.833) (Table 9). The specificity (0.758) and negative predictive value (NPV) (0.975) were only minimally lower in the cap sample. This was unexpected when considering the overall poorer outcome of MA in the cap sample. The relatively high specificity and NPV metrics could also result from the sample composition allowing for a greater margin of error for true negative outcomes, mitigating the effect of the suboptimal sample for these statistics. However, the detection prevalence (DP) was higher in the cap cohort (0.130 vs 0.106), despite the lower detection rate (DR) (0.069 vs 0.088), which illustrates the higher likelihood of an incorrect match being assigned under less ideal conditions (Table 9). This pattern becomes obvious with the F_1-Score of the sunglasses sample being notably higher (0.896) than in the cap cohort (0.625) (Table 9), suggesting that even when taking into consideration false positive and false negative outcome distribution across both samples, MA performed notably better in the sunglasses sample. Overall, the clear negative impact of a brimmed cap is shown by these results of the lenient iteration of the analysis.

In the strict iteration of the statistical analysis similar results were seen, however, a small increase in performance across both samples was noted. This better performance was particularly noticeable in the cap cohort, despite the overall decrease in the raw true positive matches (Tables 10 and 11). The most notable change, however, was the drastic decrease of false positive matches, as evidenced by the decrease in FPR for both cap (from 6.7% to 0.6%) and sunglasses (from 1.9% to 0%) samples (Table 10 and 11). This shift was identified in

conjunction with an increase in FNR in both cap (from 24.2% to 42.4%) and sunglasses (from 3.2% to 16.3%) cohorts. HA was higher in this iteration across both the cap (0.956, CI = 0.929 – 0.975) and sunglasses (0.985, CI = 0.966 - 0.995) sample, bringing both samples' accuracy at a significantly greater value than the NIR for both cohorts (Table 12). This confirmation of the HA being greater than the NIR is indicative of the strict iteration performing better in images with a greater extent of facial feature obstruction that the brimmed caps cause. The increase in HA is also noted in conjunction with an increase in CCA for both cap (from 58.0% to 68.1%) and sunglasses (from 88.4% to 90.4%) sample (Table 13), further reinforcing the better performance of the strict iteration in both samples. Despite this increase, the CCA for the cap cohort remained below the target value of 0.80 (48), expectedly as only few over half of the actual matches were scored correctly (Table 10).

However, when balanced accuracy is considered, a decrease is seen in performance of the strict iteration in both cap (from 0.845 to 0.785) and sunglasses samples (from 0.974 to 0.919) when compared to the lenient iteration. This decrease is likely an outcome of the greater amount of false negative outcomes assigned in the strict iteration, as demonstrated by the drastic decrease in sensitivity in cap (from 0.758 to 0.567) and sunglasses (from 0.968 to 0.839) samples alike (Table 13). In contrast, specificity notable increased in both samples (cap: from 0.933 to 0.994 and sunglasses: from 0.980 to 1.000). Despite this difference in specificity and sensitivity, McNemar's tests demonstrated that error proportions remained similarly distributed in the cap cohort (p-value = 0.005). In the sunglasses sample a notable tendency was seen for false negative errors (Table 12), likely due to the complete lack of false positive outcomes in this cohort (Table 11).

In conjunction with the decrease in false positive outcomes, a drastic increase in PPV was seen across both samples (cap: from 0.532 to 0.905, sunglasses: from 0.833 to 1.000) coupled with a small decrease in NPV (cap: from 0.975 to 0.959, sunglasses: from 0.997 to 0.984) (Table 13) due to the increase of false negatives in both samples (Table 10 and 11). Interestingly, the PPV increase can be expected due to the closer correspondence between the DR and DP in both the cap (DR: 0.052, DP: 0.058) and sunglasses (DR: 0.077, DP: 0.077) cohorts, despite DP and DR in both cohorts being lower than the overall prevalence (0.091) of matches within the samples (Table 13). The overall stronger performance in the sunglasses cohort is further reinforced by the F_1-Score (0.839), where despite a decrease it was still considerably higher than in the cap cohort (0.704), which in turn increased marginally within the strict iteration (Table 13).

Discussion

Facial identification is commonly used in the medicolegal setting (2,10,43) as a result of the growing amount of facial image-based evidence from CCTV surveillance systems (1). As such, testing the preferred methodology of forensic facial comparison (31) under a multitude of realistic conditions is crucial in ensuring its continued usage in courts. Arguably, the most intuitive way of criminals to attempt misleading facial comparison experts is by utilising non-conspicuous disguises when under the presence of surveillance. Despite being aware of the negative impact of disguises on recognising faces for over 40 years (18), the specific effects of these disguises on facial identification remain grossly underexamined in the actualistic setting experienced by a forensic analyst.

In a previous study (32), it was shown that forensic facial comparison can be very reliably done under good conditions, when there is a clear facial view and good quality facial images. The current study investigated the performance of the FISWG feature list for morphological analysis (MA) in actualistic conditions across CCTV recordings, but with individuals disguised by sunglasses and brimmed caps. Overall, both accuracy and reliability of MA remained within ideal ranges in the sunglasses sample (hit rate accuracy (HA): 97.9%; balanced accuracy (BA): 97.4%; chance corrected accuracy (CCA): 88.4%; reliability: $\kappa = 0.853$), suggesting that enough features are still visible to make a reliable comparison. In the cap sample, however, MA was less reliable and less accurate (HA: 91.7%; BA: 84.5%; CCA: 58.0%; reliability: $\kappa = 0.648$), indicating that wearing brimmed caps is a greater concern towards successful facial identification than sunglasses. This stronger effect of caps on effectiveness of MA is further established from the HA not being statistically higher than the no information rate of the sample in the cap cohort, indicating an overall poor predictive power in the cap subset of the sample. Incorrect matches were statistically evenly distributed between false negatives (24.2%) and false positives (6.7%) in the cap sample, while in the sunglasses sample more individual false positives were assigned, despite the low false positive rate (1.9%) and the overall lower number of errors. The design of the study should be kept in mind, where false positive rates are more likely than false negatives, as per face pool only one false negative, but nine false positives are possible. Nevertheless, it seems that false positives were a relative rarity across both sample types.

The distribution of errors was notably altered in the strict iteration, where only the higher level of confidence outcome was considered a match (43). Particularly, a noteworthy reduction in false positive outcomes across both samples (sunglasses: 0%; caps: 0.61%) was observed. This reduction in false positives was, however, coupled with a considerable increase in false negatives (sunglasses: 16.13%; caps: 42.42%), particularly in the cap sample, where almost half of the actual face matches were incorrectly categorised as not matching. While this approach is erring to the advantage of a suspect, it further highlights the notable reduction in accuracy when attempting to compare faces obscured by brimmed caps.

Comparing the outcomes of the present study to published literature is difficult as no other studies followed a feature-based systematic morphological analysis, and only a few of the studies investigating disguises in face matching reported error rates (e.g. (15,50). Most studies tested the performance of multiple observers, usually members of the public and not trained experts, without using a systematic approach (e.g. (51). Nonetheless, in their experiment involving members of the public attempting to identify the presence of a target with a brimmed cap from CCTV footage, Henderson *et al.* (15) found markedly high false positive rates, ranging between 40% and 60%. These rates were higher than their false negative rate (17%) and the false positive rates in their sample without the disguise, ranging between 10% and 37% (15). The higher error rates seen by Henderson *et al.* (15) in the cap sample is in agreement with the current study, however, the distribution of errors and high false positive rate is not. This incongruent outcome could possibly be due to overconfidence in their ability by members of the public, and not following a systematic approach.

Another study which explored the ability of a jury to identify a defendant, identified no significant difference in mean error rates between no disguise conditions (44%) and a “beanie” (hat) (41%), while identifying that dark eyeglasses resulted in the lowest mean error rate (30%) (50). Davis and Valentine (50) speculate on a number of potential factors as the cause for this inconsistency in relation to other studies. It was firstly suggested that the presence of the target in person during the matching process could provide more information than a single image would (50). This additional information and likely additional time of consideration could potentially lead to some level of familiarisation with the face, which may increase the ability to correctly identify the individual. Secondly, the fact that the participants performing the matches were recommended to use external features when comparing the dark eyeglasses disguised target may have also aided in their increased accuracy (50).

In memory tasks of facial recognition involving multiple participants from members of the public reviewing facial images, with and without eyeglasses, an increase in false alarm (false positive) rate and decrease in hit rate is reported by Terry (21), despite the actual rates not being clearly stated. A follow up study, investigating the transformation of the face by adding and removing disguises, found that the addition of eyeglasses created a specific increase in false alarms (false positives) as the face now resembled other faces with glasses, while the removal of glasses made the face easier to identify (52). This could lead to the speculation that, cognitively, eyeglasses, or other disguises might be incorporated in the facial appearance as a whole and treated as another "facial" feature. This additional feature could then be subjectively used, unsuccessfully, in facial comparison (17), which could partly explain the decrease in accuracy found with disguises in non-feature-based approaches to facial comparison. This argument should not be applicable to a systematic forensic facial comparison, which does not involve a memory task face matching test. Righi *et al.* (17) found that when eyeglasses are present during a facial comparison test, an observer would be more able to disregard them and perform the comparison without considering the disguise at all. Despite the ability to disregard eyeglasses, it is not entirely clear how they affect cognitive processes of facial matching (17). However, the consideration of glasses would appear to play a greater limiting role in holistic face review; as in a feature-based approach, like the one employed in the present study, focus is placed on each visible and distinguishable facial feature included in the feature list, which would expectedly reduce the error rate.

General agreement exists in the literature that there is a decrease in ability to recognise faces which are partially disguised by hats and eyeglasses (14,15,17,21–23,53). However, many of these studies address disguises with a focus on specific settings of passport security screening (e.g. 51,54) or impersonation (e.g. 53), often with little consistency in the actual disguise accessories used (e.g. 53). Despite this inconsistency, Noyes and Jenkins (53) found a decreased accuracy for disguised unfamiliar same or similar faces ranging from 60% to 82% as opposed to the non-disguised group with accuracies well above 90%, which are comparably similar to the chance corrected accuracies (CCA) found in this present study (58.0% to 90.4%). When compared to non-disguised faces with a similar methodological approach the accuracies of facial comparison was lower in the present study than the non-disguised faces in an ideal photographic sample (99.1%) (32). However, when compared to a similar condition CCTV sample, the sunglasses sample had a higher CCA than the non-disguised faces (82.6%) (32). Even though error rates shifted in the strict iteration of the present study, HA and CCA

increased for both sunglasses (HA: 98.5%; CCA: 90.4%) and caps (HA: 95.6%; CCA: 68.1%), while BA decreased in both samples (sunglasses: 91.9%; caps: 78.5%) likely due to the drastic increase seen in false negatives, as BA takes false negatives and false positives into consideration. The accuracy ranges identified in the present study were, for the most part, well above those identified in the literature that deals with disguises, once again reiterating the importance of a systematic approach.

The superior performance of morphological analysis using a systematic approach in the sunglasses cohort over the brimmed cap cohort is in agreement with other studies. Lee *et al.* (14) identified the lowest accuracy in their sample of head gear with a brim (57.1%) as well as Henderson *et al.* (15) who identified even lower accuracies (between 40 and 43%) in their brimmed hat face matching task. Mansour *et al.* (23) on the other hand, found sunglasses to have a more pronounced negative effect on facial recognition than hats. This was suggested to result from sunglasses obscuring internal features, namely the eyes, which are traditionally perceived as primary facial features used in identification (23). Obscuring "major" facial features could also prevent holistic processing required for face recognition (17,55), further hampering the identification process. Unfamiliar face recognition is broadly considered to rely primarily on external facial features (7). The reliance on external features could play a greater role in correct identification (7), which in turn is corroborated by the current study's findings of sunglasses having a limited effect on accuracy and error rates. Brimmed caps on the other hand, obscure a greater number of external facial features, possibly decreasing performance of MA. In contrast, Kemp *et al.* (54), when comparing face matching ability by masking external features, identified an increase in accuracy of 2.4%. Their study, however, only investigated a one-to-one possible match as opposed to both the current study and Bruce *et al.* (7) which included an array of possible match options to select from. In addition, when comparing whole faces to a face with masked external features, Kemp *et al.* (54) found a poorer performance. Despite the relevance of external and internal feature processing in facial comparison, it should be noted that the hat used by Mansour *et al.* (23) was a brimless toque. As a result, fewer overall features were obscured compared to the present study and the features obscured were primarily the more superior external features (i.e. hairline and partial forehead).

Rationally it would be expected that brimmed caps would pose a greater limitation to facial comparison than brimless caps and sunglasses, due to the greater number of facial features obscured. Sunglasses permit the review of the forehead shape, contour, and relative size as well as the details of the hairline and baldness pattern (FIG. 1 and FIG. 2), whereas the brimmed

hats often leave a shadow over the features that need to be observed (14). In the present study, the majority of the target images had facial features up to and including the superior half of the nose obstructed by the brim and the shadow it projected on the face, effectively masking both the forehead area as well as the eyes and eyebrows (FIG. 1). The sunglasses also projected a shadow, although to a lesser extent, which partially obscured the zygomatic region of the face (FIG. 2); however, greater segments of the face were visible overall (FIG. 2). As such, the overall underperformance of MA in the brimmed cap sample is likely an outcome of the combined large-scale loss of facial feature information occurring (56) and the reliance on external facial features during unfamiliar face matching (7).

Expectedly, the decrease in face recognition ability has been found to be further exacerbated when both caps and glasses are included in face recognition tasks (51). Meissner *et al.* (51) identified a notable decrease in discriminatory accuracy (a computed accuracy that accounts for false positives, but not false negatives) ranging from 57% to 73% in disguised comparisons (51). This is an example of compounding facial feature screening strategy limitations by involving both disguises that obscure features (caps and glasses). Interestingly, Abudarham and Yovel (57) reverse engineered a series of facial features of high "perceptual sensitivity", which if altered digitally resulted in altering the perception of the identity of a given face (58). These high perceptual sensitivity features were particularly consistent in allowing identification when mismatched facial poses or views were compared (57). The critical features included primarily hair colour, mouth size, eye colour, eyebrow thickness, and eye shape (57). With the exception of hair colour and mouth size, these features are located in the internal aspect of the face and most of these would be partially of fully obscured by sunglasses. While in the present study sunglasses did in fact obscure these features, so did the brimmed cap due to the positioning of the CCTV camera at a greater height than the target and at a 27° angle of incidence. Hence, while these features were obscured in both samples, the brimmed cap also obscured other features, justifiably resulting in a lower accuracy of MA. It is important to note though that these features were developed and tested on European male faces exclusively (57,58), as such they may not be applicable to the African male faces included in the present study. While male faces were selected in this study due to the higher occurrence of males as both crime offenders and victims (59,60), the same methodological approach might yield different results on female faces. As such, morphological analysis should be investigated in further studies with an additional focus on disguises more commonly employed by women,

such as scarves and cosmetic make-up, which will likely obscure different facial features to the ones discussed in this study.

In consideration of the mixed involvement of facial features and cognitive processes in facial comparison, further studies are required to investigate and comprehend this interaction within a forensic facial comparison context. It is of vital importance to confirm which features are truly critical across various population groups and morphologies and integrate that information with the various possible disguises, beyond brimmed caps and sunglasses examined in the present study. The outcomes of the present study lend themselves to recommend training observers to increase awareness of the negative effect of disguises, to allow ignoring them during the facial comparison process. This training would likely aid in increasing accuracy of forensic analysts performing identification tasks. In addition, specifically training analysts in identifying faces from internal features may also contribute to improving identification accuracy in surveillance data where the target may be disguised by hats (16).

Trained examiners have been demonstrated to have a superior performance to face recognition systems (61), where disguises can also interfere with performance. Disguises are also a prime concern in automated face detection, the initial process of semi-automated facial recognition systems (62). Recent strides have been made to automatically detect isolated parts of a face (e.g. (62,63) and previously attempts were also made to superimpose missing facial information in disguised images (64). However, until these systems are improved to permit accurate and reliable detection and recognition of faces in a manner suitable for forensic practice, morphological analysis is likely to persist as the recommended practice. Thus, addressing concerns around disguise effects and developing specific disguise based facial comparison guidelines for trained analysts are of great importance to improve morphological analysis across various types of disguises.

Conclusion

This has been the first study to test the extensive FISWG feature list for facial comparison by morphological analysis in a sample consisting of disguised faces captured by CCTV surveillance. Highly accurate and reliable performance was identified even when faces were disguised with sunglasses, while a substantial decrease in both accuracy and reliability was found in faces disguised by a brimmed cap. This was attributed to a combination of critical features being obscured by the disguises (57), the overall loss of information to consider in the

matching process (56) and the specific loss of external facial features, widely considered crucial for unfamiliar facial comparison (7).

In this context, a series of further investigations into the specific interactions between these factors are recommended, and identification practices should be adjusted in an attempt to mitigate the negative effects of the disguises. These include training forensic analysts in the specific deleterious effects of disguises to facial matching cognition (17). Training should focus on internal features of the face (16) which can further aid in obtaining the highest possible accuracy under the challenging circumstances of disguised-face surveillance evidence.

REFERENCES

1. Norris C, McCahill M, Wood D. The growth of CCTV: a global perspective on the international diffusion of video surveillance in publicly accessible space. Surveill Soc 2002;2(2/3):110–35. https://doi.org/10.24908/ss.v2i2/3.3369.

2. Vilalta CJ, Sanchez TW, Fondevila G, Ramirez M. A descriptive model of the relationship between police CCTV systems and crime. Evidence from Mexico City. Police Pract Res 2019;20(2):105–21. https://doi.org/10.1080/15614263.2018.1473770.

3. Celine J, Agustin SA. Face Recognition in CCTV Systems. 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE, 2019;111–6.

4. Jain AK, Klare B, Park U. Face matching and retrieval in forensics applications. IEEE Multimed 2012;19(1):20–20. https://doi.org/10.1109/MMUL.2012.4.

5. Burton AM, White D, McNeill A. The Glasgow face matching test. Behav Res Methods 2010;42(1):286–91. https://doi.org/10.3758/BRM.42.1.286.

6. Robertson DJ, Fysh MC, Bindemann M. Facial identity verification: five challenges facing practitioners. Keesing J Doc 2019;59(June):3–8.

7. Bruce V, Henderson Z, Greenwood K, Hancock PJB, Burton AM, Miller P. Verification of face identities from images captured on video. J Exp Psychol Appl 1999;5(4):339–60. https://doi.org/10.1037/1076-898X.5.4.339.

8. Wilkinson C, Evans R. Are facial image analysis experts any better than the general

public at identifying individuals from CCTV images? Sci Justice 2009;49(3):191–6. https://doi.org/10.1016/j.scijus.2008.10.011.

9. Bromby M. CCTV and expert evidence: addressing the reliability of new sciences. Archbold News 2006;(9):6–9.

10. Davis JP, Valentine T, Wilkinson C. Facial image comparison. In: Wilkinson C, Rynn C, editors. Craniofacial Identification. New York: Cambridge University Press, 2012;136–53.

11. Bruce V, Henderson Z, Newman C, Burton AM. Matching identities of familiar and unfamiliar faces caught on CCTV images. J Exp Psychol Appl 2001;7(3):207–18. https://doi.org/10.1037/1076-898X.7.3.207.

12. Megreya AM, Burton AM. Unfamiliar faces are not faces: evidence from a matching task. Mem Cognit 2006;34(4):865–76. https://doi.org/10.3758/BF03193433.

13. Megreya AM, Sandford A, Burton AM. Matching face images taken on the same day or months apart: the limitations of photo ID. Appl Cogn Psychol 2013;27(6):700–6. https://doi.org/10.1002/acp.2965.

14. Lee W-L, Wilkinson C, Memon A, Houston K. Matching unfamiliar faces from poor quality closed-circuit television (CCTV) footage: An evaluation of the effect of training on facial identification ability. AXIS Online J Cent Anat Hum Identif 2009;1(1):19–28.

15. Henderson Z, Bruce V, Burton AM. Matching the faces of robbers captured on video. Appl Cogn Psychol 2001;15(4):445–64. https://doi.org/10.1002/acp.718.

16. Megreya AM, Memon A, Havard C. The headscarf effect: direct evidence from the eyewitness identification paradigm. Appl Cogn Psychol 2012;26(2):308–15. https://doi.org/10.1002/acp.1826.

17. Righi G, Peissig JJ, Tarr MJ. Recognizing disguised faces. Vis cogn 2012;20(2):143–69. https://doi.org/10.1080/13506285.2012.654624.

18. Patterson KE, Baddeley AD. When face recognition fails. J Exp Psychol Hum Learn Mem 1977;3(4):406–17. https://doi.org/10.1037/0278-7393.3.4.406.

19. O'Rourke TE, Penrod SD, Cutler BL, Stuve TE. The external validity of eyewitness identification research: generalizing across subject populations. Law Hum Behav

1989;13(4):385–95. https://doi.org/10.1007/BF01056410.

20. Ryder H, Smith HMJ, Flowe HD. Estimator variables and memory for faces. In: Valentine T, Davis JP, editors. Forensic facial identification: theory and practice of identification from eyewitnesses, composites and CCTV. John Wiley & Sons, Ltd., 2015;159–83.

21. Terry RL. How wearing eyeglasses affects facial recognition. Curr Psychol 1993;12(2):151–62. https://doi.org/10.1007/BF02686820.

22. Davis JP, Tamonyte D. Masters of disguise: Super-recognisers' superior memory for concealed unfamiliar faces. 2017 Seventh International Conference on Emerging Security Technologies (EST). IEEE, 2017;44–9.

23. Mansour JK, Beaudry JL, Bertrand MI, Kalmet N, Melsom EI, Lindsay RCL. Impact of disguise on identification decisions and confidence with simultaneous and sequential lineups. Law Hum Behav 2012;36(6):513–26. https://doi.org/10.1037/h0093937.

24. Cutler BL, Penrod SD. Improving the reliability of eyewitness identification: lineup construction and presentation. J Appl Psychol 1988;73(2):281–90. https://doi.org/10.1037/0021-9010.73.2.281.

25. Ellis HD, Shepherd JW, Davies GM. Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition. Perception 1979;8(4):431–9. https://doi.org/10.1068/p080431.

26. Kramer RSS, Manesi Z, Towler A, Reynolds MG, Burton AM. Familiarity and within-person facial variability: the importance of the internal and external features. Perception 2018;47(1):3–15. https://doi.org/10.1177/0301006617725242.

27. Jenkins R, Burton AM. Stable face representations. Philos Trans R Soc B Biol Sci 2011;366(1571):1671–83. https://doi.org/10.1098/rstb.2010.0379.

28. Burton AM, Kramer RSS, Ritchie KL, Jenkins R. Identity From Variation: Representations of Faces Derived From Multiple Instances. Cogn Sci 2016;40(1):202–23. https://doi.org/10.1111/cogs.12231.

29. Megreya AM. Feature-by-feature comparison and holistic processing in unfamiliar face matching. PeerJ 2018;6:e4437. https://doi.org/10.7717/peerj.4437.

30. Towler A, White D, Kemp RI. Evaluating the feature comparison strategy for forensic face identification. J Exp Psychol Appl 2017;23(1):47–58. https://doi.org/10.1037/xap0000108.

31. Facial Identification Scientific Working Group. Facial Comparison Overview and Methodology Guidelines. 2019. Version 1.0 2019.10.25. Available at: https://fiswg.org/fiswg_facial_comparison_overview_and_methodology_guidelines_V1.0_20191025.pdf (accessed 21 January 2021)

32. Bacci N, Houlton TMR, Briers N, Steyn M. Validation of forensic facial comparison by morphological analysis in photographic and CCTV samples. Int J Legal Med 2021. https://doi.org/10.1007/s00414-021-02512-3.

33. Facial Identification Scientific Working Group. Facial Identification Scientific Working Group (FISWG) overview. 2018. Version: 2.0 2018.09.19. Available at: https://fiswg.org/FISWG_overview_v2.0_2018_09_19.pdf (accessed 21 January 2021)

34. Evison M. Forensic Facial Analysis. In: Bruinsma G, Weisburd D, editors. Encyclopedia of Criminology and Criminal Justice. New York: Springer, 2014;1713–29.

35. Houlton TMR, Steyn M. Finding Makhubu: A morphological forensic facial comparison. Forensic Sci Int 2018;285:13–20. https://doi.org/10.1016/j.forsciint.2018.01.022.

36. Megreya AM, Bindemann M. Feature instructions improve face-matching accuracy. PLoS One 2018;13(3):e0193455. https://doi.org/10.1371/journal.pone.0193455.

37. Facial Identification Scientific Working Group. Facial Image Comparison Feature List for Morphological Analysis. 2018. Version: 2.0 2018.09.11. Available at: https://fiswg.org/FISWG_Morph_Analysis_Feature_List_v2.0_20180911.pdf (accessed 21 January 2021)

38. Ritz-Timme S, Gabriel P, Obertovà Z, Boguslawski M, Mayer F, Drabik A, et al. A new atlas for the evaluation of facial features: advantages, limits, and applicability. Int J Legal Med 2011;125(2):301–6. https://doi.org/10.1007/s00414-010-0446-4.

39. Mallett X, Evison MP. Forensic Facial Comparison: Issues of Admissibility in the Development of Novel Analytical Technique. J Forensic Sci 2013;58(4):859–65.

https://doi.org/10.1111/1556-4029.12127.

40. Bacci N, Davimes J, Steyn M, Briers N. Wits Face Database. Wits Institutional Repos. Environ. Dsp. 2020. https://hdl.handle.net/10539/29924.

41. Bacci N, Davimes J, Steyn M, Briers N. Development of the Wits Face Database: an African database of high-resolution facial photographs and multimodal closed-circuit television (CCTV) recordings. F1000Research 2021. https://doi.org/10.12688/f1000research.50887.1.

42. Speckeis C. Can ACE-V be validated? J Forensic Identif 2011;61(3):201–9.

43. Steyn M, Pretorius M, Briers N, Bacci N, Johnson A, Houlton TMR. Forensic facial comparison in South Africa: State of the science. Forensic Sci Int 2018;287:190–4. https://doi.org/10.1016/j.forsciint.2018.04.006.

44. R: a language and environment for statistical computing [computer program]. MS Windows version: 3.6.3. R Core Team. R Foundation for Statistical Computing, Vienna, Austria. 2020. Available at: https://www.r-project.org

45. irr: various coefficients of interrater reliability and agreement R package. [computer program package]. Gamer M, Fellows I, Singh P, Lemon J. MS Windows version 0.84.1. 2019. Available at : https://cran.r-project.org/package=irr

46. Kuhn M. Building Predictive Models in R Using the caret Package. J Stat Softw 2008;28(5):1–26. https://doi.org/10.18637/jss.v028.i05.

47. Caret: classification and regression training R package. [computer program package]. Kuhn M. MS Windows version: 6.0-86. 2020. Available at: https://CRAN.R-project.org/package=caret

48. Watson PF, Petrie A. Method agreement analysis: A review of correct methodology. Theriogenology 2010;73(9):1167–79. https://doi.org/10.1016/j.theriogenology.2010.01.003.

49. Dietterich TG. Approximate statistical tests for comparing supervised classification learning algorithms. Neural Comput 1998;10(7):1895–923. https://doi.org/10.1162/089976698300017197.

50. Davis JP, Valentine T. CCTV on trial: Matching video images with the defendant in the

dock. Appl Cogn Psychol 2009;23(4):482–505. https://doi.org/10.1002/acp.1490.

51. Meissner CA, Susa KJ, Ross AB. Can I see your passport please? Perceptual discrimination of own- and other-race faces. Vis cogn 2013;21(9–10):1287–305. https://doi.org/10.1080/13506285.2013.832451.

52. Terry RL. Effects of Facial Transformations on Accuracy of Recognition. J Soc Psychol 1994;134(4):483–92. https://doi.org/10.1080/00224545.1994.9712199.

53. Noyes E, Jenkins R. Deliberate disguise in face identification. J Exp Psychol Appl 2019;25(2):280–90. https://doi.org/10.1037/xap0000213.

54. Kemp RI, Caon A, Howard M, Brooks KR. Improving Unfamiliar Face Matching by Masking the External Facial Features. Appl Cogn Psychol 2016;30(4):622–7. https://doi.org/10.1002/acp.3239.

55. Valentine T, Davis JP. Forensic facial identification: theory and practice of identification from eyewitnesses, composites and CCTV. Chichester, UK: John Wiley & Sons, Ltd, 2015.

56. Beaudry JL, Bullard CL, Dolin JR. Estimator Variables and Eyewitness Identification. In: Bruinsma G, Weisburd D, editors. Encyclopedia of Criminology and Criminal Justice. New York, USA: Springer, 2014;1384–94.

57. Abudarham N, Yovel G. Reverse engineering the face space: Discovering the critical features for face identification. J Vis 2016;16(3):40. https://doi.org/10.1167/16.3.40.

58. Abudarham N, Shkiller L, Yovel G. Critical features for face recognition. Cognition 2019;182(April 2018):73–83. https://doi.org/10.1016/j.cognition.2018.09.002.

59. Farley R. Homicide Trends in the United States. Demography 1980;17(2):177. https://doi.org/10.2307/2061058.

60. Maluleke R. Crime Statistics Series Volume V: Crime against Women in South Africa. 2018 Jun. Report No.: 03-40-05 . Available at: http://www.statssa.gov.za/publications/Report-03-40-05/Report-03-40-05June2018.pdf

61. White D, Dunn JD, Schmid AC, Kemp RI. Error Rates in Users of Automatic Face Recognition Software. PLoS One 2015;10(10):e0139827. https://doi.org/10.1371/journal.pone.0139827.

62. Zorin A, Abramov N. Disguised Face Detection. In: Kim KJ, Kim H-Y, editors. Information Science and Applications. Springer, 2020;509–17.

63. Kumaar S, Vishwanath RM, Omkar SN, Majeedi A, Dogra A. Disguised Facial Recognition Using Neural Networks. 2018 IEEE 3rd International Conference on Signal and Image Processing (ICSIP). IEEE, 2018;28–32.

64. Yoshino M, Noguchi K, Atsuchi M, Kubota S, Imaizumi K, Thomas CDL, et al. Individual identification of disguised faces by morphometrical matching. Forensic Sci Int 2002;127(1–2):97–103. https://doi.org/10.1016/S0379-0738(02)00115-9.

Tables

Table 1. Analysis trials face pools: cohort subsample breakdown.

Analysis Cohort	Test Trials	No-match Trials	Analysis Trials	Repeat Trials	Total Trials
Cap CCTV to ST photos	33	4	37	5	42
Sunglasses CCTV to ST photos	31	3	34	5	39

ST = standard, CCTV = Closed Circuit Television

Table 2. Facial pose distribution throughout face pool sample.

Analysis Cohort	Anterior View	Left 45° View	Left Lateral View	Right 45° View	Right Lateral View	Total Trials
Cap CCTV to ST photos	7	10	7	8	6	42
Sunglasses CCTV to ST photos	6	8	7	6	9	39

ST = standard, CCTV = Closed Circuit Television

Table 3. FISWG facial component list for facial comparison by morphological analysis (37).

Number	Facial Component
1	Skin (texture, colour, tone etc.)
2	Face/Head Outline (contour)
3	Face/Head Composition (proportions)
4	Hairline/Baldness Pattern (detailed shape)
5	Forehead (relative size and contour)
6	Eyebrows (angle, shape, relative size and hair details)
7	Eyes (colour, position, angle, relative size etc.)
8	Cheeks (prominence and dimples)
9	Nose (shape, position, symmetry, relative size etc.)
10	Ears (shape, position, protrusion and relative size)
11	Mouth (angulation, shape, symmetry, relative size etc.)
12	Chin/Jawline (shape, relative size, symmetry, prominence, and definition)

13	Neck (relative size, prominence, folds, veins etc.)
14	Facial Hair (presence, distribution, length, style, hair details etc.)
15	Facial Lines (creases, wrinkles etc.)
16	Scars (location, shape, orientation, relative size, tonality, depth etc.)
17	Facial Marks (moles, discoloration patches, freckles etc.)
18	Alterations (tattoos, piercings etc.)

Table 4. Contingency table reference for confusion matrix of positive and negative matches adapted from Kuhn (46)**.**

	True face match	*True face exclusion*
Matched by analyst	True Positive (TP)	False Positive (FP)
Excluded by analyst	False Negative (FN)	True Negative (TN)

Table 5. Tabulation of statistical metrics used based on the Caret R package (46,47)**, equations are available as supplementary information.**

Statistic	**Definition**	**Equation**
Sensitivity	Also called true positive rate or recall, is defined as the measure of how effectively a test can perform correct matches	1
Specificity	Also called true negative rate, is a measure of how effective a test can correctly discard a match	2
Prevalence	The total number of truly correct matches in the entire cohort sample	3
Positive predictive value (PPV)	Also called Precision, this metric indicates the probability of a trial scored as a match being truly a correct match by calculating the ratio of correctly assigned matches to total positive matches made by the analyst	4
Negative predictive value (NPV)	This statistic is the probability of a trial scored as an exclusion being truly a correct exclusion	5
Detection rate	The proportion of the cohort sample correctly matched.	6

	Under perfect circumstances, the detection rate would be identical or as close to the detection prevalence value as possible, which is in the range of 9 to 10% in this study due to the organisation of the sample.	
Detection prevalence	The proportion of trials scored as matches by the analyst	7
Balanced Accuracy	An adjusted accuracy calculation that considers the unbalanced distribution of the sample and is calculated by including sensitivity and specificity.	8
F1-Score	A weighted (harmonic) mean of Precision and Sensitivity including both false positive and false negative rates. To obtain an F-Score the weight of Precision and Sensitivity in the equation is adjusted according to the requirements of the study; equal weighting was given to both variables in our study. In the resulting equation 10, β is assigned a value of 1 and it is termed an F1-Score.	9

Table 6. Contingency table for confusion matrix of the lenient iteration of the Cap CCTV cohort. In this table the analysed sample is contrasted to the ideal perfect matches and exclusions to demonstrate performance of the facial comparisons.

	Actual face match	*Actual face exclusion*		*Actual face match*	*Actual face exclusion*
Matched by analyst	TP = 25 (TPR = 75.8%)	FP = 22 (FPR = 6.7%)	*Ideal perfect match*	TP = 33	FP = 0
Excluded by analyst	FN = 8 (FNR = 24.2%)	TN = 307 (TNR = 93.3%)	*Ideal perfect exclusions*	FN = 0	TN = 329

TP = true positive; FP = false positive; TN = true negative; FN = false negative; TPR = true positive rate; FPR = false positive rate; FNR = false negative rate; TNR = true negative rate.

Table 7. Contingency table reference for confusion matrix of the lenient iteration of the Sunglasses CCTV cohort analysed. In this table the analysed sample is contrasted to the ideal perfect matches and exclusions to demonstrate performance of the facial comparisons.

	Actual face match	*Actual face exclusion*		*Actual face match*	*Actual face exclusion*
Matched by analyst	TP = 30 (TPR = 96.8%)	FP = 6 (FPR = 1.9%)	*Ideal perfect match*	TP = 31	FP = 0
Excluded by analyst	FN = 1 (FNR= 3.2%)	TN = 303 (TNR = 98.1%)	*Ideal perfect exclusions*	FN = 0	TN = 309

TP = true positive; FP = false positive; TN = true negative; FN = false negative; TPR = true positive rate; FPR = false positive rate; FNR = false negative rate; TNR = true negative rate.

Table 8. Confusion matrix function results and chance corrected accuracies across cohorts of analysis.

	Cap CCTV	*Sunglasses CCTV*
Hit rate accuracy	0.917	0.979
95% confidence interval	0.884 - 0.943	0.958 - 0.992
No information rate (NIR)	0.909	0.909
P-Value (Acc>NIR)	0.331	< 0.001
McNemar's Test P-Value	0.018	0.131
Chance Corrected Accuracy (CCA)	58.0%	88.4%
P-value (for CCA)	< 0.001	< 0.001

Table 9. Tabulation of the lenient iteration predictive quality values calculated across cohorts of analyses.

	Cap CCTV	*Sunglasses CCTV*
Sensitivity (Recall)	0.758	0.968

Specificity	0.933	0.980
Prevalence	0.091	0.091
Positive predictive value (PPV)	0.532	0.833
Negative predictive value (NPV)	0.975	0.997
Detection Rate (DR)	0.069	0.088
Detection Prevalence (DP)	0.130	0.106
Balanced Accuracy	0.845	0.974
F_1-Score	0.625	0.896

Table 10. Contingency table reference for confusion matrix of the strict iteration of the Cap CCTV cohort. In this table the analysed sample is contrasted to the ideal perfect matches and exclusions to demonstrate performance of the facial comparisons.

	Actual face match	*Actual face exclusion*		*Actual face match*	*Actual face exclusion*
Matched by analyst	TP = 19 (TPR = 57.6%)	FP = 2 (FPR = 0.6%)	*Ideal perfect match*	TP = 33	FP = 0
Excluded by analyst	FN = 14 (FNR = 42.4%)	TN = 327 (TNR = 99.4%)	*Ideal perfect exclusions*	FN = 0	TN = 329

TP = true positive; FP = false positive; TN = true negative; FN = false negative; TPR = true positive rate; FPR = false positive rate; FNR = false negative rate; TNR = true negative rate.

Table 11. Contingency table reference for confusion matrix of the strict iteration of the Sunglasses CCTV cohort analysed. In this table the analysed sample is contrasted to the ideal perfect matches and exclusions to demonstrate performance of the facial comparisons.

	Actual face match	*Actual face exclusion*		*Actual face match*	*Actual face exclusion*
Matched by analyst	TP = 26 (TPR = 83.9%)	FP = 0 (FPR = 0%)	*Ideal perfect match*	TP = 31	FP = 0

Excluded by analyst	FN = 5 (FNR = 16.1%)	TN = 309 (TNR = 100%)	*Ideal perfect exclusions*	FN = 0	TN = 309

TP = true positive; FP = false positive; TN = true negative; FN = false negative; TPR = true positive rate; FPR = false positive rate; FNR = false negative rate; TNR = true negative rate.

Table 12. Confusion matrix function results across the strict iteration of each cohort of analysis.

	Cap CCTV	*Sunglasses CCTV*
Hit rate accuracy	0.956	0.985
95% confidence interval	0.929 – 0.975	0.966 - 0.995
No information rate (NIR)	0.909	0.909
P-Value (Acc>NIR)	< 0.001	< 0.001
McNemar's Test P-Value	= 0.005	= 0.074
Chance Corrected Accuracy (CCA)	68.1%	90.4%
P-value (for CCA)	< 0.001	< 0.001

Table 13. Tabulation of predictive quality values calculated across the strict iteration of each cohort of analysis.

	Cap CCTV	*Sunglasses CCTV*
Sensitivity (Recall)	0.576	0.839
Specificity	0.994	1.000
Prevalence	0.091	0.091
Positive predictive value (PPV)	0.905	1.000
Negative predictive value (NPV)	0.959	0.984
Detection Rate (DR)	0.052	0.077
Detection Prevalence (DP)	0.058	0.077
Balanced Accuracy	0.785	0.919
F_1-Score	0.704	0.839

Figures

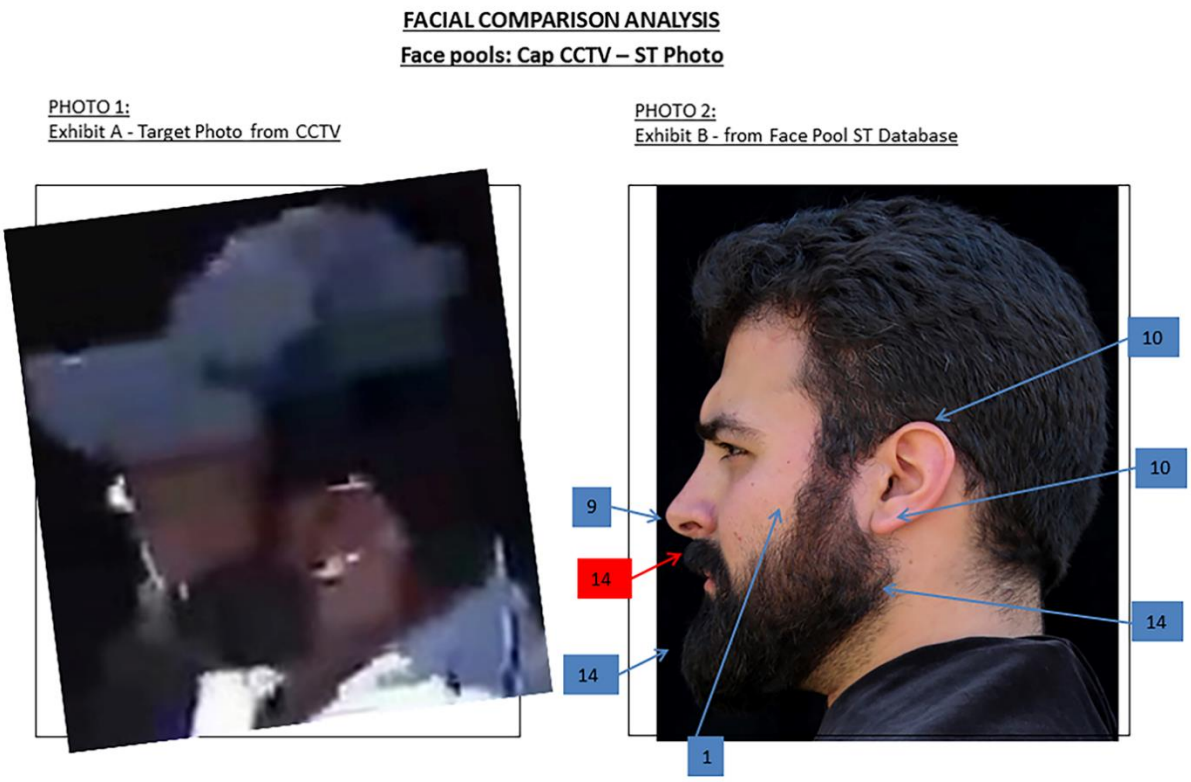


FIG. 1 - *Facial comparison analysis example from the cap obstruction CCTV sample.* *A target image from CCTV was placed next to each possible match. Standardised (ST) photo (right) was compared to the target image from the CCTV recordings (left), using the FISWG facial component list (37). Similar facial components were highlighted by facial component number, in blue. If present, dissimilar components were marked in red.*

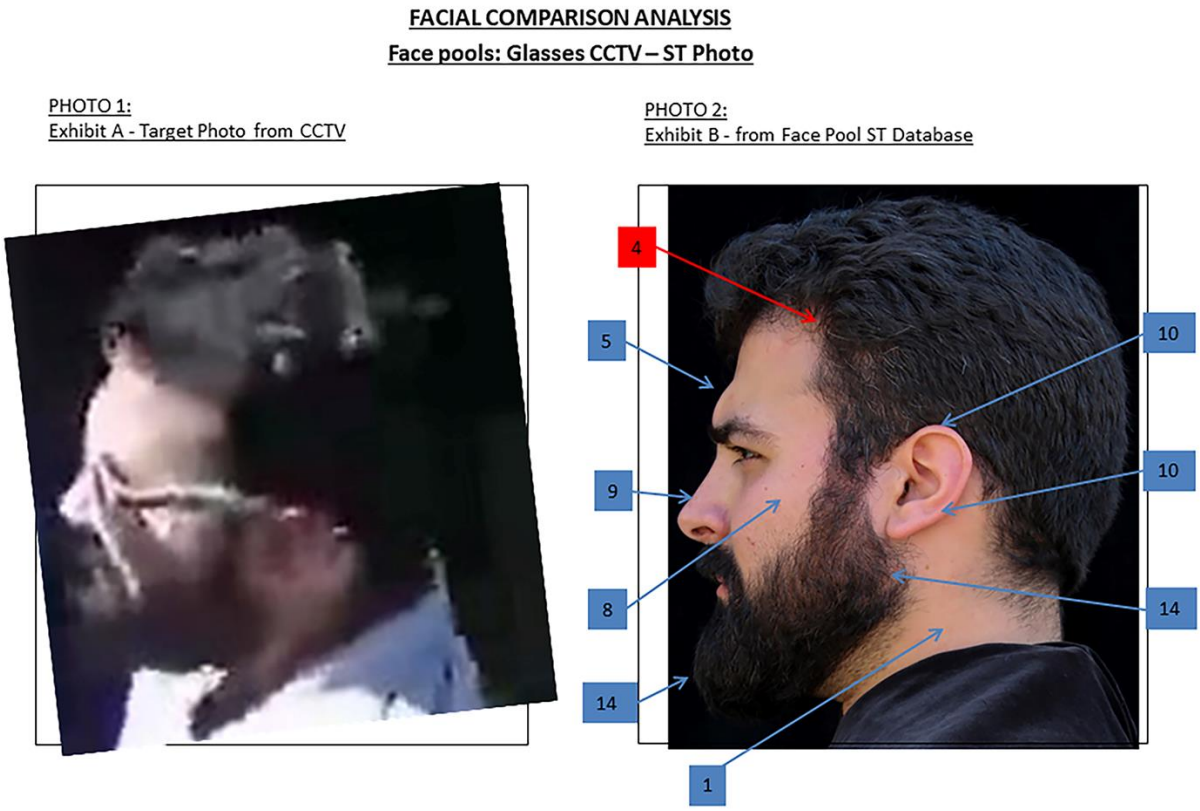


FIG. 2 - *Facial comparison analysis example from the glasses obstruction CCTV sample.* *A target image from CCTV was placed next to each possible match. Standardised (ST) photo (right) was compared to the target image from the CCTV recordings (left), using the FISWG facial component list (37). Similar (blue) and dissimilar (red) facial components were highlighted by facial component number.*



FIG. 3 - ***Bar graphs of the facial comparison outcomes under a lenient (A) and strict (B) iteration within the cap CCTV cohort by trial.*** *Each bar represents the 10 comparisons in each face pool trial, with individual segments representing the type of outcome as follows: TP = true positive; FP = false positive; TN = true negative; FN = false negative; IA = inconclusive analysis*

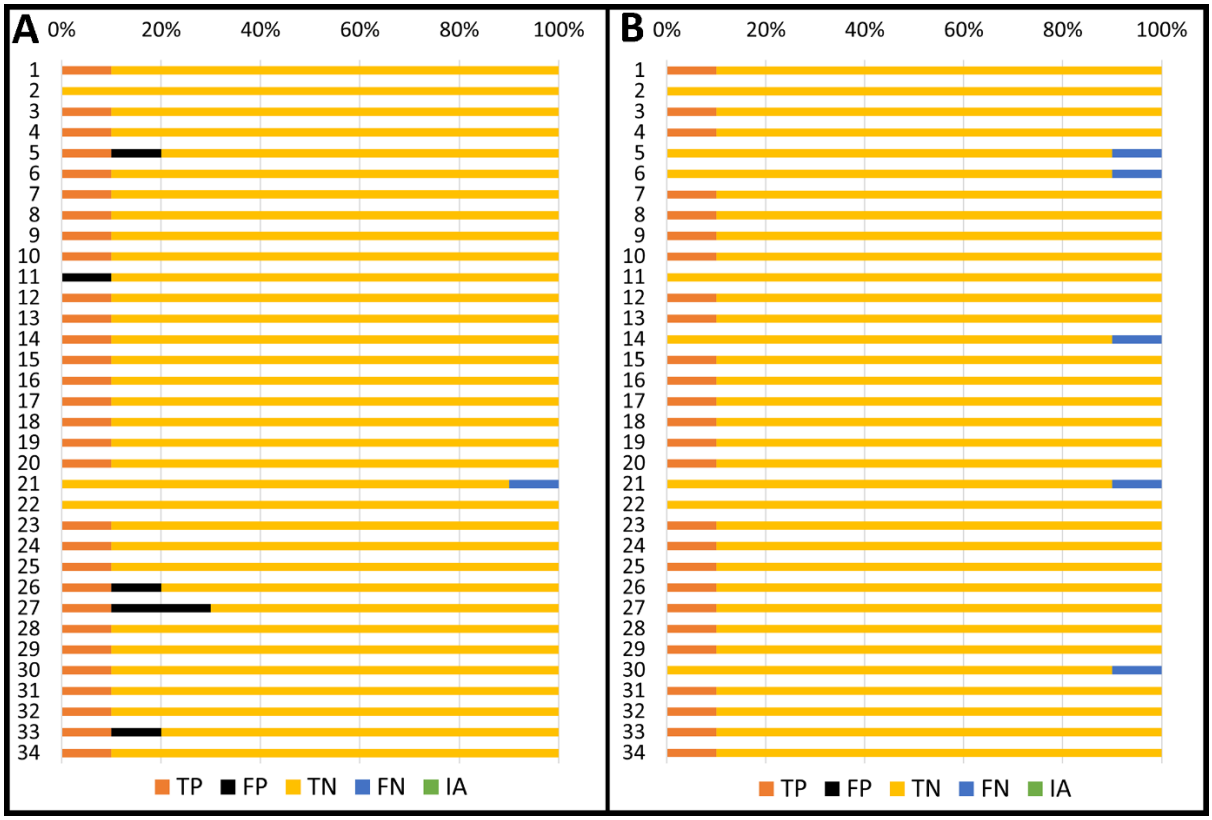


FIG. 4 - ***Bar graphs of the facial comparison outcomes under a lenient (A) and strict (B) iteration within the sunglasses CCTV cohort by trial.*** *Each bar represents the 10 comparisons in each face pool trial, with individual segments representing the type of outcome as follows: TP = true positive; FP = false positive; TN = true negative; FN = false negative; IA = inconclusive analysis*

6. Concluding discussion

Validation is arguably the most important aspect of any scientific forensic practice. Scientists have often been approached by representatives of the legal system to develop forensic methods or technologies or report on evidence based on their relevant scientific expertise. There is an expectation both from the public and the legal system of well-documented scientific foundations, reliable data, and rigorous logical application being employed by forensic experts (Illes *et al.*, 2020). However, this may not always be the case as history and research has demonstrated (Peterson and Markham, 1995; Garrett and Neufeld, 2009).

While a "general acceptance" of morphological analysis as a valid method for facial comparison exists among the scientific community (Steyn *et al.*, 2018; Facial Identification Scientific Working Group, 2019), a lack of scientific and rigorous testing has been specifically noted in the field of facial identification (Mallett and Evison, 2013). While validation is important, the legal systems should not be placed at a disadvantage by disregarding generally accepted methods in circumstances where methods may be novel and only employed on a case-by-case basis (Mallett and Evison, 2013). The greater problem arises when this acceptance and precedent is relied upon for the continued use of untested forensic practices. To this end, the current project was able to provide some long-needed validation testing of forensic facial comparison *via* morphological analysis, while also developing the first large-scale multi-modal African facial image database. This is now the largest African database of CCTV recordings and high-resolution facial photographs available. The infrastructure used to collect the CCTV data most closely resembles standard CCTV installations in a real-world setting across the various conditions reflected in the cohorts of analysis used in the study (e.g., analogue, and internet protocol cameras). Recruitment of participants was purposive and based on the selection of male individuals with facial morphology of phenotypically black South African descent. Convenience recruitment was also employed by targeting the large student population on campus and as such, participant age was only confirmed for the purposes of consent and specific age was not collected initially. With the majority of the participants being students at the University, the database is composed primarily of young adult individuals. While the author is currently involved in a large-scale process to obtain self-reported ages from participants these data are not yet available for inclusion. An additional note regarding the database composition involves the intentional lack of inclusion of siblings or twins, to the best of the author's knowledge as the level of similarity would require an additional database subdivision too timeous to collect.

6.1 Morphological analysis validation

Morphological analysis on data derived from this newly developed database (Manuscript 1; Chapter 2: (Bacci, *et al.*, 2021) using the FISWG feature list was found highly accurate and reliable when comparing optimal standardised photographs to wildtype unstandardised photographs. In optimal photographs an almost perfect chance corrected accuracy (99.1%) was found. When comparing average digital CCTV recordings to photographs (Manuscript 2; Chapter 3: (Bacci, *et al.*, 2021), while still considered acceptable at 82.6%, a decreased accuracy and reliability was immediately noted in CCTV images. This decreased performance in CCTV was primarily attributed to the variation of conditions intrinsic in comparing varied types of image data obtained under different conditions and by different equipment. Specifically, quality of images was poorer when extracted from the average CCTV system, due to a variety of reasons. Firstly, the actual resolution of the average digital CCTV camera was notably lower (4MP) than the photographic cameras utilised to compile the photographic sample (18MP). The CCTV field of view was also much larger, partly due to the face-to-objective distance of approximately 3000 mm. Therefore, a larger area was recorded with this lower resolution, effectively making the resolution of a captured face even smaller. Secondly, the angle of incidence of 27° formed between the CCTV camera used in this study and the captured face appeared to limit the visible portions of the face to a certain extent, and possibly altering relative facial feature proportions. This change in perspective and the challenges it created for facial comparison likely contributed to the lower accuracy and reliability identified in the CCTV sample, when compared to the photographic sample, despite no additional head tilt being included in the facial images in this study. In addition, the lighting was notably different between photographs and CCTV recordings, which made facial characteristics that rely on colour (i.e., skin tone and colour) particularly inconsistent even between confirmed matches. Additional arguments can be made about the effects of lighting variations causing overexposure of certain features, effectively obscuring them. However, these concerns extend beyond the scope of the current study and should be investigated independently. Notwithstanding these challenges, the almost perfect accuracies and the low false positive rates identified (< 1.6%), particularly when considering a more stringent iteration of the analysis, encourages the use of morphological analysis in a legal context from both optimal photographs and average CCTV installations, as it is preferable to limit false convictions.

6.2 Facial comparison across CCTV conditions

While average digital CCTV camera performance was good, this study considered performance across various CCTV camera conditions (Manuscript 3; Chapter 4: (Bacci *et al.*, 2021). The eye-level digital CCTV data was found to have better results than standard CCTV installation, while analogue CCTV footage was found to limit MA performance notably relative to digital recordings. Decreasing the camera to face distance (800 mm) as well as an effective 0° angle of incidence appears to have contributed to eye-level camera-derived images being closer to ideal for facial comparison. In fact, eye-level digital CCTV recording-based facial comparison had a similar accuracy to the optimal photograph only sample. While this may indicate the importance of angle of incidence and distance, to fully confirm the extent of the influence of these factors on facial comparison, further independent testing would be required.

The effects of distance and the facial distortion generated have been previously investigated in the context of craniofacial superimposition (Stephan, 2015, 2017). Stephan (2015) identified that discrepancy in objective-to-face/skull distances between photographs to be assessed using superimposition can result in perspective distortion of facial features. This effect was found to be more pronounced at shorter distances, particularly below 6000 mm (Stephan, 2015). The effect of distance-based perspective distortion would be a crucial consideration in methods that rely on metric information (e.g. superimposition and photo-anthropometry), as at distances below 1000 mm even a discrepancy of 100 mm in objective to face distance can yield a perspective distortion greater than 1% (Stephan, 2015). While perspective distortion could possibly affect morphological comparison of facial features, the qualitative approach of morphological analysis and the large number of listed features would be likely to mitigate small levels of perspective distortion between the compared images. However, to confirm this, further research into comparing the perspective distortion and its specific effect in otherwise fully controlled images would be necessary.

Despite these uncertainties, from the results of the current study, a recommendation is made that CCTV systems should shift towards the use of high-definition cameras installed at eye-level heights. However, this would limit the cost-effectiveness of CCTV installations as one camera would have a more limited field of view at the lower height (Damjanovski, 2014). As such, more cameras would need to be installed to cover areas previously covered by a single or pair of cameras (Damjanovski, 2014). A less considered challenge of CCTV installations at lower heights is vandalism. Cameras installed at arm's reach are significantly easier to sabotage

or vandalise in a criminal scenario, rendering them useless and increasing future costs of replacing/repairing the equipment. While tamper-proof installations and decoy options are available, they by default increase the CCTV system's cost. These options are commercially available yet when left up to the consumer's discretion may be overlooked on the side of cost-effectiveness until a specific CCTV installation fails to aid in crime resolution. CCTV installations should be directed to follow these specifications to produce images better suited to facial examination.

The overall lower quality of analogue CCTV images appeared to also play a significant role in effective use of MA. While this sample was intended to simulate a worst-case scenario of poor-quality monochrome surveillance data, it included three factors that increased the difficulty of facial comparison, namely a pronounced angle of incidence (22°), lack of colour, and particularly low-resolution images. Isolating which of these factors impacted MA performance the most was not possible by design. In combination these three factors were highly disruptive to the facial comparison process. The detrimental effect of the lack of colour by itself is questionable. Many facial examiners conduct their comparisons in greyscale with the consideration that colour can be considered misleading. The possibility of colour being misleading was also observed in the present study when attempting to match skin colour between images. However, it is clear that all three factors notably inhibited application of MA the most out of all CCTV conditions included in the study - even more so than the inclusion of disguises. In fact, the analogue sample was the only sample to perform worse when a strict iteration was considered in the scoring criteria. Accuracy in this analogue sample was as low as of 33.1% with an extremely high rate of false negatives (75.2%). This outcome suggests that under particularly suboptimal conditions, even a strict iteration of including only the highest level of confidence for positive matches is inefficient. In turn, the decrease in the strict iteration may be worth considering as a marker for suboptimal conditions. It should perhaps be deliberated that in future studies when a strict iteration results in lower performance of MA in a sample, that sample be considered too subpar to use accurately in facial comparison. As a result, particularly in cases where no factors can be used to exclude an individual as a match, recordings that perform poorly under a strict iteration should not be utilised for positive identification. However, since there is a high prevalence of subpar image data available for police services' use (Gibelli *et al.*, 2016) the images should not be completely discarded from evidence, as they may still provide useful information, but not for forensic facial comparison. This follows the recommendations by the European Network of Forensic Science Institutes

suggesting triaging based on the quality of images to ensure beneficial application of facial comparison and effectiveness in caseload management (ENFSI, 2018).

In the analogue sample comparison, it was also noted that the best measure of accuracy from the various ones computed was chance corrected accuracy using a weighted Cohen's kappa. Chance corrected accuracy appeared to vary the most and was considered the most representative when compared to simple hit rate or raw accuracy and balanced accuracy, which included weighting based on specificity and sensitivity. The main flaw of these two accuracies was attributed to sample organisation which skewed towards a particular score (true negatives) with an inherent higher likelihood to be scored as such. Therefore, the non-chance corrected accuracies appeared deceptively higher despite other measures of test performance indicating a more complex outcome. As such, in future studies concerning facial comparison, these similar statistics should be considered with a particular focus on chance corrected accuracy. In addition, conducting varied threshold analyses with strict and lenient iterations can provide further insights and better applications for the legal system while attempting to maintain the inviolability of presumption of innocence. Lastly, the aforementioned skewness of the sample, which resulted from the study's design that was intended to simulate a situation where multiple potential matches are available to a facial examiner in South Africa, and may not reflect current practices in many actual circumstances where only 1:1 may be more common. The large volume of comparisons resulting from these 1:10 comparisons also necessitated the use of holistic review of the faces prior to detailed morphological analysis due to time constraints, as each morphological analysis comparison is intended to take over two hours according to the FISWG guidelines (Facial Identification Scientific Working Group, 2012). While this process was undertaken in the most conservative way possible by allocating any face with even the slightest similarity to the target image as a potential match, it is still not best practice in facial examination. Repeating the study with a series of 1:1 comparisons using strictly morphological analysis of evenly distributed matches and non-matches may demonstrate superior statistical performance of morphological analysis and would avoid the predominance of non-matches among the results.

6.3 Effects of disguises on facial comparison

Beyond the optimisation of surveillance system installations for facial identification purposes, disguises persist as a limiting factor in facial comparison. As such, the last part of the study investigated the effect of disguised faces in facial comparison (Manuscript 4; Chapter 5: Bacci

et al., 2021). A decreased efficiency was seen when attempting to identify faces disguised by brimmed caps. These faces were recorded with the angle of incidence of the standard CCTV camera and the natural lighting from the sun, which created shadows over the face as a result of the cap's brim. The shadows and angle of capture resulted in extensive facial feature obstruction and information loss. This effectively rendered comparison much harder as almost only the lower half of the face and the ears could be contrasted.

Contrary to the effect of brimmed caps, a better performance was found when faces were disguised by sunglasses, even when compared to standard digital CCTV conditions without sunglasses (Manuscript 2; Chapter 3: Bacci, *et al.*, 2021), yet not better than eye-level CCTV (Manuscript 3; Chapter 4: Bacci *et al.*, 2021). While this is a peculiar effect of sunglasses, it has been observed by another study testing live subject to image identification (Davis and Valentine, 2009). One of their explanations to the increased ability of participants to recognise a face disguised by sunglasses was the guidance provided to participants performing face matching tasks to rely on the external facial features of subjects disguised by sunglasses (Davis and Valentine, 2009). This group of facial features is in fact considered a principal focus when performing unfamiliar face matching and is corroborated by studies that demonstrate the importance of external facial features when performing facial comparisons (Bruce *et al.*, 1999; Kemp *et al.*, 2016).

Loss of facial feature information might also occur in relation to the use of headscarves, often worn by many female members of Islamic communities. As Islam is the fastest growing religion across the world (Pew Research Center Religion & Public Life, 2015), and many countries, such as South Africa (GHS, 2017), have well-established Muslim communities (Pew Research Center Religion & Public Life, 2015) it can be important to consider the effect of headscarves in facial comparison. Megreya *et al.* (2012) found that Middle Eastern observers had an advantage in correctly matching faces from internal facial features due to the "headscarf effect". This effect involves the overreliance on internal facial features during unfamiliar facial processing as a result of cultural influences in Middle Eastern populations, where members of the public wear headscarves routinely (Megreya *et al.*, 2012). This overreliance may have an unexpected negative impact when Middle Eastern observers are tasked with comparing faces disguised by eyeglasses, which specifically obscure internal facial features.

The success of the present study's application of morphological analysis was attributed, in great part, to the FISWG feature list used (Facial Identification Scientific Working Group,

2018). It has been previously demonstrated that any form of feature instructions or feature lists raise the performance of facial comparison tasks (Towler *et al.*, 2017; Megreya, 2018; Megreya and Bindemann, 2018), more so across trained analysts (Towler *et al.*, 2017). The present study strongly reinforces that outcome, and further encourages the use of feature lists in morphological analysis applications. An often underdiscussed consideration of using a systematic approach and a feature list involves the option to discard potential dissimilarities. This can be done by considering what conditions may be rationally affecting the visibility, presentation, and variation of a certain facial feature to exclude its specific dissimilarity under the distorting conditions (e.g., lighting differences made the skin tone and colour appear distinct among two otherwise corresponding images), as was done in the present study. This benefit is possible due to the large volume of features to be considered. If no detailed consideration was being insistently given to each facial feature listed and available there is the risk of excluding of a positive face match simply due to one feature varying under different image conditions. The exact number of minimum features required is not clearly stated in the feature list (Facial Identification Scientific Working Group, 2018) or the concluding statements (Steyn *et al.*, 2018) and should be investigated in future studies. In addition, investigating specifically which facial features are most reliable and the extent to which each of them contribute to the process of facial comparison by morphological analysis, can aid in developing criteria for comparison allowance under different disguised or obstructed faces. Once established, these criteria could be reinforced by training analysts on which features to prioritise under each type of disguise. This approach could also be useful in situations where facial features are not visible due to data loss or any other physical obstructions. This could prove particularly useful as the forms of "acceptable disguises" change throughout time. For example, since 2020, with the spread of COVID-19, face masks have become commonplace. These items of personal protective equipment have been found to decrease facial recognition algorithm performance by 5% to 50% depending on the specific algorithm and extent of the face covered (Ngan *et al.*, 2020). These effects also varied based on the colour and shape of the masks (Ngan *et al.*, 2020). It would hence be crucial to consider face masks in future studies comparing morphological analysis across disguised faces to confirm the severity of the negative impact on human observer-based facial comparison.

6.4 Limitations

The aim of this study was to test morphological analysis broadly and extensively across a variety of circumstances. However, it was impossible to cover all possible permutations of

conditions in which image-based data may be presented for forensic facial identification. As such, the recommendations and outcomes of the presents study should be considered under these limitations. The overarching origin for most of these limitations stems from the organisation of the study. The study attempted to simulate real-world conditions, with its array of limitations, in a select number of scenarios. These scenarios included ideal comparable photographic images, average digital CCTV, eye-level installation digital CCTV, average monochrome analogue CCTV, and two common disguises, sunglasses, and brimmed caps. The varied circumstances of facial data were pre-set under certain conditions to attempt some level of standardisation required for experimentation. In doing so, although realistic, the conditions were limited to the major questions broadly investigated by each manuscript of the overall study. Although these broad categories were considered as realistic examples of CCTV image quality and conditions, there are multiple factors that influence the quality of an image for facial identification. Image quality relies on more than just equipment resolution capacity; it involves lighting conditions, angle of incidence, objective-to-face distance, distortions, colour, visibility of features and more. In testing the specific conditions outlined in each manuscript, controlling for or identifying which of the multiple factors that contributed to these results was impossible. Even when within likely tolerable degrees, these factors cannot in good faith be isolated from one another in the current study design. Future investigations should approach each of these factors under more controlled circumstances with single factor variations. For example, a secondary set of face pools could be developed from the analogue CCTV data set of the Wits Face Database (Manuscript 1; Chapter 2: Bacci, *et al.*, 2021) without converting the data into monochrome images. In this manner the actual effects of colour absence during the facial comparison process can be investigated in a low-resolution analogue sample.

Lighting is also a known factor in image quality, yet in the present study opting for realistic conditions all images were captured during the day and with ample amounts of light. Most current CCTV cameras can function at night as well. These “night-vision” or low light cameras emit infrared (IR) radiation with an emittance range of approximately 20 meters for most midrange cameras (Damjanovski, 2014). This IR radiation is then reflected onto a surface and received by a lens with an IR to visible light converter (Surette, 2005; Kruegle, 2007). These IR images, however, are only recorded in monochrome and as such, pose similar challenges as the traditional black and white CCTV, while also creating further image alterations in a multitude of ways due to insufficiencies of the converter (Kruegle, 2007). Furthermore, typical forms of optical distortion found in normal cameras may be more pronounced in IR cameras

due to the longer wavelength of IR radiation (Kruegle, 2007). This type of footage is hard to obtain in an experimental setting due to the requirement of capture conditions being nearly completely dark. Testing of morphological analysis should be conducted on a sample of IR surveillance data of comparable quality and conditions to the standard average CCTV camera used in the present study to ascertain whether the effect of IR image capture negatively impacts facial comparison.

The reliance on equipment and systems not specifically procured or installed for this study resulted in data loss and corruption, which extended data capture timelines and limited the overall data available to conduct the study. This limitation is a crucial consideration when planning the development or expansion of face databases. Updating databases is vital to adequate testing, particularly when considering that the current study only investigated the faces of South African males of African descent. While this is a good step towards expanding the diversity in populations groups included in face databases, not testing morphological analysis and the FISWG feature list on female individuals is a notable limitation to be aware of. In principle, the descriptive approach of the feature list is subjective and generic enough that it should be applicable invariably across sex and population groups. However, minor nuances in face variations may contribute to fluctuation in accuracies and reliability based on the observers' biases and abilities. It is well documented that a within-group face recognition advantage (previously called own- or cross-race bias) exists (Shapiro and Penrod, 1986; Bothwell *et al.*, 1989; Anthony *et al.*, 1992; Meissner and Brigham, 2001; Chiroro *et al.*, 2008; Kokje *et al.*, 2018). This effect has been documented to extend beyond the antiquated and contentious concept of "race". A study demonstrated that South Africans of European descent had an advantage in facial recognition for other South Africans of European descent but not for North Americans of European descent (Chiroro *et al.*, 2008). Similarly, South Africans of African descent were at an advantage with other South Africans of African descent but not with North Americans of African descent (Chiroro *et al.*, 2008). A similar effect has also been identified for age (Wright and Stroud, 2002) and sex (Rehnman and Herlitz, 2006). The regional focus of this within-group effect is primarily attributed to contact and perceptual experience with a specific group of individuals (Brigham and Malpass, 1985; Meissner and Brigham, 2001; Chiroro *et al.*, 2008). The fact that this effect persists and varies through age groups as well (Wright and Stroud, 2002), would likely suggest that varying the exposure of an analysts to these different groups of people would provide perceptual expertise in facial recognition across population, age and sex. This is an important consideration in light of the

analyst that performed the facial comparisons in the current study being of European descent and originally from Europe. However, the primary analyst has been permanently in South Africa for more than 10 years, allowing ample time and ability to familiarise themselves with the South African faces of various descents, sexes, and ages. Even in the circumstance that within-group advantage may have negatively skewed the results of these studies, the overall outcomes remain positive in a within-group situation, as the accuracy and reliability would likely be expected to be higher than presented when investigating a group for which perceptual expertise exists for the analyst. On the other hand, the use of the extensive FISWG list may reduce the within-group effect and help to place facial comparison on a firm, non-biased foundation irrespective of analysts' perceptual expertise.

6.5 Future directions

Validating morphological analysis according to the FISWG guidelines (Facial Identification Scientific Working Group, 2018, 2019), is a crucial stepping stone to rendering forensic facial comparison admissible in court. However, to truly understand its practical strengths and limitations an array of further studies is required. With the present studies and its recommendations for future investigations around facial comparison, more stringent requirements for CCTV installations and the morphological analysis process can be developed as minimum requirements for facial identification analysis. To continue this work, the Wits Face Database will need to be updated and expanded to include many more possible permutations of analysis, including female individuals, cosmetic make-up treatments, and face mask disguises. Future studies should also attempt to quantify the acceptable loss of facial feature information in order to successfully compare faces across a multitude of possible situations. Once that goal is achieved, better guidelines and practice frameworks can be created for legal procedures involving forensic facial comparison by morphological analysis.

While the present study tested two specific CCTV camera specifications under a few condition and installation variations, there is an immense number of manufacturers that produce CCTV equipment with different specifications, requirements, and support. Testing the extent to which various market standard cameras can affect facial comparison would be an ideal goal to strive towards. However, before attempting such a level of fine-tuning of facial comparison practice and requirements, broader aspects should be investigated. These would include investigating the contributions of each of the various aspects that seemed to decrease morphological analysis performance in the current study, particularly in an attempt to determine an empirical threshold

for suitable image quality. To do this the common factors described above that affect quality should be investigated. For instance, developing a thorough understanding of distance related distortion effects on morphological analysis between images from CCTV cameras and photographs could generate awareness of which features are altered more notably and hence increase analysis at unfavourable distances. This is an important consideration for future work due to the varied conditions most CCTV systems are installed and tailored to. While the alternative of comparing faces captured under the exact same conditions would likely be more effective, it may not be feasible or cost-effective outside of an experimental scenario. In addition, the time disparity may introduce further limitations on the equipment and conditions of image capture (e.g., different lighting, damaged camera etc.). In addition, studying the precise effect of camera angle of incidence on morphological analysis in isolation would also contribute to improving its application. Clear thresholds for determining the angle steepness that significantly inhibits facial comparison will aid in screening utility of current image data and to guide future surveillance system installation planning. Incorporating the average head tilt in these investigations would further contribute to perfecting these standards beyond the experimental scenario.

Actual digital image quality and minimum resolution allowing for facial comparison to take place should also be investigated. Based on this study's conclusions with regard to low-resolution analogue CCTV, further investigations are needed in order to define a clear, quantifiable lower-end threshold that permits analysis. This will aid both the surveillance industry and the forensic analysts conducting analyses. A useful consideration for future studies investigating all aspects of image quality in facial comparison would be to use an image quality scoring system. An example of such a scale was presented by Schüler and Obertová (Schüler and Obertová, 2020). Implementing this scale in conjunction to the FISWG feature list for morphological analysis could aid in identifying a threshold of confidence for the analysis process based on image quality.

The benefits of using a feature list, and the FISWG list specifically, are now evident and obvious. During this study, some slight shortcomings were noted, and an alternative/adjusted feature list classification scheme could be considered. An in-depth classification system extending beyond the subjective description and observer-perceived similarity used in current morphological analysis could prove successful. The FISWG feature list (tabulated in full in Appendix B) has extensively outlined individual aspects of facial features to focus on during the analysis process and provided some descriptor examples that can aid in report and

conclusion writing for law enforcement (Facial Identification Scientific Working Group, 2018). However, some of the descriptors can be superficial and vague in their outline. An example of this would be with skin tone terms inclusive of luminance and colour with no distinction or definitions of the two or with the hairline shape and various facial contours including no further detailed terminology or descriptors. While this is effective for a feature list utilised as an instruction set, providing further information on the potential morphologies of the various individual descriptors could aid in limiting some of the subjectivity of morphological analysis. Past attempts at classification schemes have been shown to be error prone (e.g. Vanezis *et al.*, 1996; Ritz-Timme *et al.*, 2011a). These classification systems, however, employed simplistic often single-worded classifications and relied on a maximum of 43 facial components (Ritz-Timme *et al.*, 2011a). In doing so it would be easy to obtain similar scores for multiple faces due to the low permutations of classifications while also at risk of misclassifying poorly defined component classes. An adjusted feature classification scheme should be based on the extensive FISWG list, while creating well-defined classifications of each trait that can be categorised and tested in a semi-quantitative manner. Such a feature list should be compiled by incorporating classifications previously made in relevant anatomy, medicine, or anthropology literature, when available, and based on the facial physiognomy of available datasets across various population groups to ensure it can be applied in multiple countries. In using an alternative approach, this morphological scoring system could be applied to both the target and the database photographs independently and no direct comparison would take place in this stage of the analysis. Once the scores are complete for each image, they would then be contrasted to each other. From these scoring values of each facial component and its characteristics a composite score of similarity could then potentially be obtained and allow a level of quantification of the similarity or dissimilarity of a given face. Such an approach remains to be tested in future studies.

6.6 Concluding remarks

The present thesis described an extensive test of morphological analysis and the FISWG feature list across varied conditions of facial CCTV images and photographs. The performance of morphological analysis was found accurate and valid, particularly when conditions were optimal (i.e., high-resolution photographs, high resolution CCTV with limited perspective distortion/angle of incidence and with no disguises or sunglasses). Among the suboptimal conditions, image quality appeared to have the most notable effect on facial comparison performance (analogue CCTV recordings), while brimmed caps were found to be the second

most interfering circumstance. Across both of these samples, the primary limiting factor appeared to be overall loss of facial feature information, with caps obscuring almost half of the face and the poor quality of the analogue video making most of the facial details indistinguishable.

There are a multitude of vulnerabilities present in most surveillance system vendor equipment that have in the past led to endangering safety, interfering with police and law-enforcement and pose theft risks (Costin, 2016). In addition, average humans were found to outperform automated recognition algorithms under suboptimal conditions (O'Toole *et al.,* 2012), which are the predominant type of data quality that criminal investigations rely on (Gibelli *et al.,* 2016). The alternative of using super-recognisers, individuals with exceptional ability to process and recognise faces (Russell *et al.,* 2009) has also been recommended (Davis and Tamonyte, 2017). However, explicitly relying on a select group of people with innate superior facial recognition ability is not a feasible strategy in most circumstances, and trained experts should be considered the gold standard as a more structured and consistent option for legal investigations. As such, with these concerns in mind and considering the positive results of the present study, it is reinforced that trained human observer-based morphological analysis should persist as the primary approach to facial comparison for identification purposes as recommended by FISWG (Facial Identification Scientific Working Group, 2019). While this study did not control for every possible condition variation, morphological analysis with the FISWG feature list proved successful in an array of realistic circumstances. Specific condition variations should be considered in future studies to attempt fine tuning the method's accuracy and reliability in each of specific scenario.

References - Concluding discussion

Anthony, T., Copper, C. and Mullen, B. (1992) 'Cross-Racial Facial Identification: A Social Cognitive Integration', Personality and Social Psychology Bulletin, 18(3), pp. 296–301. doi: 10.1177/0146167292183005.

Bacci, N., Davimes, J., et al. (2021) 'Development of the Wits Face Database: an African database of high-resolution facial photographs and multimodal closed-circuit television (CCTV) recordings', F1000Research, 10, p. 131. doi: 10.12688/f1000research.50887.1.

Bacci, N., Houlton, T. M. R., et al. (2021) 'Validation of forensic facial comparison by morphological analysis in photographic and CCTV samples', International Journal of Legal Medicine. doi: 10.1007/s00414-021-02512-3.

Bothwell, R. K., Brigham, J. C. and Malpass, R. S. (1989) 'Cross-Racial Identification', Personality and Social Psychology Bulletin, 15(1), pp. 19–25. doi: 10.1177/0146167289151002.

Brigham, J. C. and Malpass, R. S. (1985) 'The Role of Experience and Contact in the Recognition of Faces of Own- and Other-Race Persons', Journal of Social Issues, 41(3), pp. 139–155. doi: 10.1111/j.1540-4560.1985.tb01133.x.

Bruce, V. et al. (1999) 'Verification of face identities from images captured on video.', Journal of Experimental Psychology: Applied, 5(4), pp. 339–360. doi: 10.1037/1076-898X.5.4.339.

Chiroro, P. M. et al. (2008) 'Recognizing faces across continents: The effect of within-race variations on the own-race bias in face recognition', Psychonomic Bulletin and Review, 15(6), pp. 1089–1092. doi: 10.3758/PBR.15.6.1089.

Costin, A. (2016) 'Security of CCTV and Video Surveillance Systems', in Proceedings of the 6th International Workshop on Trustworthy Embedded Devices. New York, NY, USA: ACM, pp. 45–54. doi: 10.1145/2995289.2995290.

Damjanovski, V. (2014) CCTV from light to pixels. 3rd edn, CCTV from light to pixels. 3rd edn. Edited by V. Damjanovski. Oxford: Elsevier.

Davis, J. P. and Tamonyte, D. (2017) 'Masters of disguise: Super-recognisers' superior memory for concealed unfamiliar faces', in 2017 Seventh International Conference on Emerging Security Technologies (EST). IEEE, pp. 44–49. doi: 10.1109/EST.2017.8090397.

Davis, J. P. and Valentine, T. (2009) 'CCTV on trial: Matching video images with the defendant in the dock', Applied Cognitive Psychology, 23(4), pp. 482–505. doi: 10.1002/acp.1490.

ENFSI (2018) *Best Practice Manual for Facial Image Comparison*. Available at: https://enfsi.eu/wp-content/uploads/2017/06/ENFSI-BPM-DI-01.pdf (Accessed: 9 June 2021).

Facial Identification Scientific Working Group (2012) FISWG Guidelines for Facial Comparison Methods. Version 1.0 - 2012-02-02. Available at: https://www.fiswg.org/FISWG_GuidelinesforFacialComparisonMethods_v1.0_2012_02_02.pdf (Accessed: 16 March 2021).

Facial Identification Scientific Working Group (2018) Facial Image Comparison Feature List for Morphological Analysis. Version 2.0 2018.09.11. Available at: https://fiswg.org/FISWG_Morph_Analysis_Feature_List_v2.0_20180911.pdf (Accessed on 16 March 2021).

Facial Identification Scientific Working Group (2019) Facial Comparison Overview and Methodology Guidelines. Version 1.0 2019.10.25. Available at: https://fiswg.org/fiswg_facial_comparison_overview_and_methodology_guidelines_V1.0_20191025.pdf (Accessed on 16 March 2021).

Garrett, B. L. and Neufeld, P. J. (2009) 'Invalid Forensic Science Testimony and Wrongful Convictions', Virginia Law Review, 95(1), pp. 1–97.

GHS (2017) 'General Household Survey, South Africa', Statistics South Africa, P0318(May), p. 195.

Gibelli, D. et al. (2016) 'The identification of living persons on images: A literature review', Legal Medicine, 19, pp. 52–60. doi: 10.1016/j.legalmed.2016.02.001.

Illes, M., Wilson, P. and Bruce, C. (2020) 'Forensic epistemology: A need for research and pedagogy', Forensic Science International: Synergy, 2, pp. 51–59. doi: 10.1016/j.fsisyn.2019.11.004.

Kemp, R. I. et al. (2016) 'Improving Unfamiliar Face Matching by Masking the External Facial Features', Applied Cognitive Psychology, 30(4), pp. 622–627. doi: 10.1002/acp.3239.

Kokje, E., Bindemann, M. and Megreya, A. M. (2018) 'Cross-race correlations in the abilities to match unfamiliar faces', Acta Psychologica, 185(December 2017), pp. 13–21. doi:

10.1016/j.actpsy.2018.01.006.

Kruegle, H. (2007) CCTV surveillance: analog and digital video practices and technology. Second. Edited by H. Kruegle. Burlington, MA, USA: Elsevier Butterworth–Heinemann.

Mallett, X. and Evison, M. P. (2013) 'Forensic Facial Comparison: Issues of Admissibility in the Development of Novel Analytical Technique', Journal of Forensic Sciences, 58(4), pp. 859–865. doi: 10.1111/1556-4029.12127.

Megreya, A. M. (2018) 'Feature-by-feature comparison and holistic processing in unfamiliar face matching', PeerJ, 6, p. e4437. doi: 10.7717/peerj.4437.

Megreya, A. M. and Bindemann, M. (2018) 'Feature instructions improve face-matching accuracy', PLOS ONE. Edited by P. J. Hills, 13(3), p. e0193455. doi: 10.1371/journal.pone.0193455.

Megreya, A. M., Memon, A. and Havard, C. (2012) 'The Headscarf Effect: Direct Evidence from the Eyewitness Identification Paradigm', Applied Cognitive Psychology, 26(2), pp. 308–315. doi: 10.1002/acp.1826.

Meissner, C. A. and Brigham, J. C. (2001) 'Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review.', Psychology, Public Policy, and Law, 7(1), pp. 3–35. doi: 10.1037/1076-8971.7.1.3.

Ngan, M., Grother, P. and Hanaoka, K. (2020) Ongoing Face Recognition Vendor Test (FRVT) part 6A: Face recognition accuracy with masks using pre-COVID-19 algorithms. Gaithersburg, MD. doi: 10.6028/NIST.IR.8311.

O'Toole, A. J. et al. (2012) 'Comparing face recognition algorithms to humans on challenging tasks', ACM Transactions on Applied Perception, 9(4), pp. 1–13. doi: 10.1145/2355598.2355599.

Peterson, J. L. and Markham, P. N. (1995) 'Crime Laboratory Proficiency Testing Results, 1978–1991, II: Resolving Questions of Common Origin', Journal of Forensic Sciences, 40(6), p. 13871J. doi: 10.1520/JFS13871J.

Pew Research Center Religion & Public Life (2015) The Future of World Religions: Population Growth Projections, 2010-2050. Available at: http://www.pewforum.org/2015/04/02/religious-projections-2010-2050/. (Accessed on 16 March 2021).

Rehnman, J. and Herlitz, A. (2006) 'Higher face recognition ability in girls: Magnified by own-sex and own-ethnicity bias', Memory, 14(3), pp. 289–296. doi: 10.1080/09658210500233581.

Ritz-Timme, S. et al. (2011) 'A new atlas for the evaluation of facial features: advantages, limits, and applicability', International Journal of Legal Medicine, 125(2), pp. 301–306. doi: 10.1007/s00414-010-0446-4.

Russell, R., Duchaine, B. and Nakayama, K. (2009) 'Super-recognizers: People with extraordinary face recognition ability', Psychonomic Bulletin & Review, 16(2), pp. 252–257. doi: 10.3758/PBR.16.2.252.

Schüler, G. and Obertová, Z. (2020) 'Visual identification of persons: facial image comparison and morphological comparative analysis', in Obertová, Z., Stewart, A., and Cattaneo, C. (eds) Statistics and Probability in Forensic Anthropology. First. London, UK: Elsevier Academic Press, pp. 313–330. doi: 10.1016/B978-0-12-815764-0.00025-3.

Shapiro, P. N. and Penrod, S. (1986) 'Meta-analysis of facial identification studies.', Psychological Bulletin, 100(2), pp. 139–156. doi: 10.1037/0033-2909.100.2.139.

Stephan, C. N. (2015) 'Perspective distortion in craniofacial superimposition: Logarithmic decay curves mapped mathematically and by practical experiment', Forensic Science International, 257, pp. 520.e1-520.e8. doi: 10.1016/j.forsciint.2015.09.009.

Stephan, C. N. (2017) 'Estimating the Skull-to-Camera Distance from Facial Photographs for Craniofacial Superimposition', Journal of Forensic Sciences, 62(4), pp. 850–860. doi: 10.1111/1556-4029.13353.

Steyn, M. et al. (2018) 'Forensic facial comparison in South Africa: State of the science', Forensic Science International, 287, pp. 190–194. doi: 10.1016/j.forsciint.2018.04.006.

Surette, R. (2005) 'The thinking eye', Policing: An International Journal of Police Strategies & Management, 28(1), pp. 152–173. doi: 10.1108/13639510510581039.

Towler, A., White, D. and Kemp, R. I. (2017) 'Evaluating the feature comparison strategy for forensic face identification.', Journal of Experimental Psychology: Applied, 23(1), pp. 47–58. doi: 10.1037/xap0000108.

Vanezis, P. et al. (1996) 'Morphological Classification of Facial Features in Adult Caucasian Males Based on an Assessment of Photographs of 50 Subjects', Journal of

Forensic Sciences, 41(5), p. 13998J. doi: 10.1520/jfs13998j.

Wright, D. B. and Stroud, J. N. (2002) Age Differences in Lineup Identification Accuracy: People Are Better with Their Own Age, Law and Human Behavior. 26(6), pp. 641–654. doi: 10.1023/A:1020981501383

Appendix A – Ethics Clearance Certificate

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

R14/49 Mr Nicholas Bacci et al

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M171026

NAME: (Principal Investigator)	Mr Nicholas Bacci et al
DEPARTMENT:	School of Anatomical Sciences Matrix - East Campus, University of the Witwatersrand
PROJECT TITLE:	Reliability of Forensic Facial Identification from CCTV Recordings with Implications for Admissibility to Court
DATE CONSIDERED:	27/10/2017
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	Dr Nanette Briers
APPROVED BY:	Professor CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	13/03/2018

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary, Third Floor, Faculty of Health Sciences, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown, 2193, University of the Witwatersrand. I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. **I agree to submit a yearly progress report**. The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. In this case, the study was initially reviewed in October and will therefore be due in the month of October each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature

Date 14/03/2018

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

Appendix B -Tabulated Facial Comparison Feature List (FISWG, 2018)

Component Characteristics	Characteristic Descriptors
1 - Skin	
1.1 Overall Skin Appearance	• Overall Texture • Overall Tone: e.g. luminance and colour
2 – Face/head Outline	
2.1 Shape of Cranial Vault	• Portrait Contour • Profile Contour
2.2 Overall Shape of Face	• Portrait Contour • Profile Contour
3 – Face/head Composition	
3.1 Proportions/Position of Features on Face	• Approximate width of nose relative to eye distances (e.g., interpupillary distance, individual eye width, or overall distance between outer corners) • Approximate width of mouth relative to eye distances • Approximate width of nose relative to mouth • Approximate distance from nose to upper lip relative to face length • Approximate distance from chin to lower lip relative to face length • Ear position relative to eyes, nose, and mouth • Eye position relative to face length
4 - Hairline	
4.1 Forehead Hairline	• Detailed Shape (e.g. symmetry, “widow’s peak”, part line, “cowlick”)
4.2 Hairline Right Side 4.3 Hairline Left Side	• Detailed Shape
4.4 Cranial Baldness Pattern	• Detailed Shape
5 - Forehead	
5.1 Forehead Shape	• Relative height

	• Relative width • Slope/contour (visible in profile)
5.2 Brow Ridges	• Prominence • Continuity
6 - Eyebrows	
6.1 Right Eyebrow 6.2 Left Eyebrow	• Shape (may include details) • Size (width and length of eyebrow relative to eye size) • Lateral eyebrow vertical end position relative to medial eyebrow vertical position (tilt of eyebrow) • Vertical end position of lateral eyebrow relative to the lateral canthus • Vertical end position of medial eyebrow relative to the medial • Horizontal end position of lateral eyebrow relative to lateral canthus • Horizontal end position of medial eyebrow relative to medial canthus • Conjoined left-right eyebrows ("unibrow") • Density of hair within eyebrow and distribution of density • Hair details (e.g. texture, length, thickness, shape, colour) • Noticeably longer hairs
6.3 Asymmetry Between Right and Left Eyebrows	• e.g. overall shape, size, position, and hair details
7 - Eyes	
7.1 Inter-Eye Distance	• Distance between right and left eye
7.2 Right Eye Fissure Opening (Outline) 7.3 Left Eye Fissure Opening (Outline)	• Shape • Angle
7.4 Right Upper Eyelid 7.5 Left Upper Eyelid (Including Lashes)	• Prominence (e.g. visibility, folds) • Protrusion

	• Visibility of the crease above the upper eyelid (superior palpebral furrow) • Position in relation to iris and/or pupil • Lash characteristics (e.g. length, density, flow, irregular)
7.6 Right Lower Eyelid 7.7 Left Lower Eyelid (Including Lashes)	• Prominence (e.g. visibility, folds) • Protrusion • Visibility of the crease below the lower eyelid (inferior palpebral furrow) • Visibility of infraorbital furrow (a place where a line or wrinkle may appear parallel to and below the lower eyelid running from near the inner canthus and following cheek bone laterally) • Position in relation to iris and/or pupil • Lash characteristics (e.g. length, density, flow, irregular)
7.8 Right Eyeball Prominence 7.9 Left Eyeball Prominence	• Degree of Protrusion
7.10 Right Eye Sclera 7.11 Left Eye Sclera	• Visibility of blood vessels and defects • Colour
7.12 Right Iris 7.13 Left Iris	• Colour • Visibility • Diameter relative to eye opening • Position relative to eye opening (in front view) • Irregularity in pupil
7.14 Right Eye Medial Canthus 7.15 Left Eye Medial Canthus	• Caruncle (fleshy prominence at inner eye corner) • Shape and angle of inner corner of the eye
7.16 Right Eye Lateral Canthus 7.17 Left Eye Lateral Canthus	• Shape and angle of outer corner of the eye
7.18 Asymmetry Between Right and Left Eyes	• Shape, Angle • Off-set (e.g. one eye higher than the other)

	• Eyelids (e.g. one drooping, one retracted) & Eyelashes • Colour • Iris and pupil position (i.e. cross-eyed) • e.g. overall shape, size, position, and hair details
8 - Cheeks	
8.1 Right Cheekbone 8.2 Left Cheekbone	• Prominence
8.3 Right Cheek Shape 8.4 Left Cheek Shape	• Presence of dimple
9 - Nose	
9.1 Nasal Outline (Profile and Front view)	• Overall Shape • Length and/or width relative to rest of face • Prominence • Symmetry
9.2 Nasal Root (Bridge)	• *Front View: w*idth, length, shape, depth • *Profile View:* length, depth, angle
9.3 Nasal Body	• *Front View: w*idth, length, shape, angle • *Profile View:* length, angle, contour
9.4 Nasal Tip	• Shape (in front and profile view) • Angle (e.g. up, down) • Symmetry
9.5 Nasal Base	• Width • Height • Deviation to the right or left
9.6 Nasal Base: Alae (Wings of Nose)	• Thickness • Symmetry • Shape
9.7 Nasal Base: Nostrils (Nasal Openings)	• Shape and size of opening • Symmetry • Hair
9.8 Nasal Base: Columella (Soft Tissue between Nostrils)	• Width and length • Relative position

	• Symmetry
10 - Ears	
10.1 Asymmetry Between Left and Right Ears	• Size • Shape • Protrusion • Positioning (e.g. One is above the • other)
10.2 Right Ear Protrusion 10.3 Left Ear Protrusion	• Extent of Protrusion
10.4 Overall Right Ear 10.5 Overall Left Ear	• Size • Shape • Angle
10.6 Right Ear Helix—Superior/Inferior (Tail) 10.7 Left Ear Helix—Superior/Inferior (Tail)	• Size • Shape
10.8 Right Ear Tubercles (Auricular Tubercle) 10.9 Left Ear Tubercles (Auricular Tubercle)	• Size • Shape • Quantity
10.10 Right Ear Antihelix 10.11 Left Ear Antihelix	• Size • Shape
10.12 Right Ear Crura of Antihelix (Superior, Inferior) 10.13 Left Ear Crura of Antihelix (Superior, Inferior)	• Size • Shape
10.14 Right Ear Triangular Fossa 10.15 Left Ear Triangular Fossa	• Size • Shape
10.16 Right Ear Crus of Helix 10.17 Left Ear Crus of Helix	• Size • Shape
10.18 Right Ear Scaphoid Fossa 10.19 Left Ear Scaphoid Fossa	• Size • Shape
10.20 Right Ear Concha (Superior, Inferior) 10.21 Left Ear Concha (Superior, Inferior)	• Size • Shape
10.22 Right Ear Tragus 10.23 Left Ear Tragus	• Size • Shape
10.24 Right Ear Antitragus 10.25 Left Ear Antitragus	• Size • Shape

10.26 Right Ear Intertragic Notch 10.27 Left Ear Intertragic Notch	• Size • Shape
10.28 Right Ear Anterior Knob 10.29 Left Ear Anterior Knob	• Size • Shape
10.30 Right Ear Anterior Notch 10.31 Left Ear Anterior Notch	• Size • Shape
10.32 Right Ear Posterior Auricular Furrow 10.33 Left Ear Posterior Auricular Furrow	• Size • Shape
10.34 Right Ear Lobule (Lobe) 10.35 Left Ear Lobule (Lobe)	• Size • Shape
10.36 Ear Abnormalities	• *e.g. Cleft lobe; cauliflower*
11 - Mouth	
11.1 Philtrum	• Prominence • Width of ridges • Width of furrow • Symmetry
11.2 Overall Mouth	• Shape • Symmetry
11.3 Upper Lip	• Shape • Fullness • Protrusion • Symmetry • Upper Vermilion Border Shape (e.g. “Cupid's bow”) and definition • Detail (e.g. tubercle, lip creases, alae)
11.4 Lower Lip	• Shape • Fullness • Protrusion • Symmetry • Lower Vermilion Border Shape & definition • Detail (e.g. median sulcus, tori, lip creases)
11.5 Lip Fissure (Opening between Lips)	• Shape • Symmetry

	• Degree of contact/occlusion along length of opening • Corners/angles of mouth (Labial Commissure)
11.6 Mouth Asymmetry	• Difference between left and right sides
11.7 Overall Dental Occlusion (Contact between Upper and Lower Teeth)	• Symmetry • Degree of contact/occlusion
11.8 Prognathism (Forward protrusion of mouth complex)	• Degree
11.9 Characteristic Detail of Teeth	• Shape • Size • Alignment/Position (e.g. gaps, crooked, missing) • Condition (e.g. wear, damage, disease, colour)
11.10 Mouth Abnormalities	• *E.g.* Cleft Lip
12 – Chin/Jawline	
12.1 Chin (Profile and Frontal view)	• Overall Shape • Length and/or width relative to rest of face • Prominence • Symmetry • Details (e.g. cleft, dimple, mental groove; refer to wrinkles and folds)
12.2 Jawline (from Chin to Gonial Angle)	• Shape • Definition (e.g. jowls)
12.3 Gonial Angle (Angle of the Jaw)	• Shape • Definition
13 – Neck	
13.1 Neck (Overall)	• Width • Height • Details (e.g. musculature, veins, • wrinkles, folds, wattle, “double chin”)
13.2 Laryngeal Prominence (Adam's Apple)	• Shape • Size

	• Prominence • Location on neck •
14 – Facial Hair	
14.1 Facial Hair Above Upper Lip 14.2 Facial Hair Below Lower Lip	• Shape/Spatial distribution (including overall hair length) • Texture • Symmetry • Density and distribution of density including gaps • Variation in colour/tonality • Orientation (slanted, straight) • Outline/edge definition (e.g., sharp, irregular) • Continuity with facial hair on side(s) or below/above mouth • Noticeably Longer Hairs
14.3 Facial Hair on Right Side 14.4 Facial Hair on Left Side	• Shape/Spatial distribution (including overall hair length) • Texture • Symmetry • Density and distribution of density including gaps • Variation in colour/tonality • Orientation (slanted, straight) • Outline/edge definition (e.g., sharp, irregular) • Continuity with facial hair on side(s) or below/above mouth • Noticeably Longer Hairs
14.5 Facial Hair on Neck, below Chin/Jawline	• Shape/Spatial distribution (including overall hair length) • Texture • Symmetry

	• Density and distribution of density including gaps • Variation in colour/tonality • Orientation (slanted, straight) • Outline/edge definition (e.g., sharp, irregular) • Continuity with facial hair on side(s) or below/above mouth • Noticeably Longer Hairs
15 – Facial Lines	
15.1 Forehead Creases (Wrinkles)	• Distribution • Orientation (vertical or horizontal) • Quantity • Pattern (Including relation to one another) • Depth/Prominence
15.2 Vertical Glabellar Line(s)	• Length • Pattern (including relation to one another) • Depth/Prominence
15.3 Nasion Crease	• Distribution • Quantity • Pattern (Including relation to one another) • Depth/Prominence
15.4 Bifid Nose Crease	• Depth/Prominence • Length
15.5 Wrinkles Adjacent to Right Eye (Crow's Feet) 15.6 Wrinkles Adjacent to Left Eye (Crow's Feet)	• Distribution • Quantity • Pattern (Including relation to one another) • Depth/Prominence
15.7 Right Superior Palpebral Crease 15.8 Left Superior Palpebral Crease (Crease between the Upper Eyelid and the	• Visibility • Position • Depth/Prominence

Top of the Bony Orbit)	• Shape
15.9 Right Lower Lid Crease 15.10 Left Lower Lid Crease (Crease between the Lower Eyelid and the Bottom of the Bony Orbit)	• Visibility • Position • Depth/Prominence • Shape
15.9 Right Lower Lid Crease 15.10 Left Lower Lid Crease (Crease between the Lower Eyelid and the Bottom of the Bony Orbit)	• Distribution • Quantity • Pattern (Including relation to one another) • Depth/Prominence
15.11 Right Infraorbital Creases 15.12 Left Infraorbital Creases (Creases below the Eyes)	• Distribution • Quantity • Pattern (Including relation to one another) • Depth/Prominence
15.13 Upper Lip Creases (Vertical Wrinkles above Upper Lip)	• Distribution • Quantity • Pattern (Including relation to one another) • Depth/Prominence
15.14 Lower Lip Creases (Vertical Wrinkles below Lower Lip)	• Distribution • Quantity • Pattern (Including relation to one another) • Depth/Prominence
15.15 Mentolabial Sulcus (Horizontal Crease or Fold between Lower Lip and Chin)	• Shape • Length • Depth/Prominence
15.16 Right Nasolabial Crease/Folds 15.17 Left Nasolabial Crease/Folds (Creases or Folds extending from Nose to Corners of Mouth)	• Distribution • Quantity • Pattern (Including relation to one another) • Depth/Prominence
15.18 Right Marionette Lines 15.19 Left Marionette Lines	• Pattern • Depth/Prominence

15.20 Cleft Chin	• Depth/Prominence • Size
15.21 Right Buccal Creases/Folds 15.22 Left Buccal Creases/Folds (Cheek to Chin)	• Distribution • Quantity • Pattern (Including relation to one another) • Depth/Prominence
15.23 Wrinkles on Neck	• Distribution • Quantity • Pattern (Including relation to one another) • Depth/Prominence
16 - Scars	
16.1 Scars	• Location • Shape • Orientation • Size • Colour/Tonality • Depth/Prominence
17 – Facial Marks	
17.1 Skin Marks (e.g. freckles, moles, acne, rosacea, birth marks, bruises, abrasions, vitaligo, dark/light patches, etc.)	• Location/Distribution (Including relation to one another) • Shape • Size • Colour • Prominence
18 - Alterations	
18.1 Piercing	• Location • Description
18.2 Makeup	• Location • Description (Shape, Size, Colour, etc.)
18.3 Tattoo (Including Cosmetic)	• Location • Description (Content, Shape, Size, Colour, etc.)

Appendix C – Supplementary Material Manuscript 1

PARTICIPANT INFORMATION SHEET – FACIAL PHOTOGRAPHS AND CCTV RECORDINGS

STUDY TITLE: *Reliability of forensic facial identification from CCTV recordings with implications for admissibility to court.*

Good day, my name is Nicholas Bacci and I am currently enrolled for a PhD degree at the University of the Witwatersrand. As part of the degree I am carrying out a research project on forensic facial image comparison. For this project, I will study ways to improve the process of forensic facial identification, especially from video obtained from closed circuit television (CCTV) systems. I will be comparing different parts of the face between photos and CCTV footage for similarities. This method is called morphological analysis. Although this method is very often used by the police, little is known on how reliable it is. The main possible benefit of this project is that it will help to correctly match a face to another face on different images. The results of the study will greatly assist in criminal investigations and prosecution as well as missing persons' identification. The study will also prevent wrongful imprisonment of innocent people. I would kindly like to invite you to participate in this study.

PURPOSE OF THE STUDY

Photos of different faces and CCTV footage of the same faces will be used to compare the similarities and dissimilarities between the two sets of images. From this comparison, an analyst will decide whether the faces are a match or not. The aim of this process is to assess if the analyst is able to correctly identify the photo matching the CCTV footage from a random pool of photos. You are invited to participate as one of the photographed individuals.

PARTICIPANT INVOLVEMENT

If you decide to participate in this study, I will compare your CCTV recordings to your photos. I will do this by comparing the best matching (in terms of angle and position) faces to one of the photos that will be taken of you in one trial. Each trial will include a CCTV image to be compared to a pool of ten photos. Only one of them will be your photo. Nine extra photos will be added from social media based on people resembling your face. I will then try to match your CCTV recording to your photo in this pool of faces. I will only take

five photos of your head and face on the university premises, and record CCTV footage as you walk past the set of CCTV cameras. You will also be asked to wear a jacket hiding your clothes. The process will not take longer than ten minutes. Your participation in this study is voluntary. Refusal to participate will involve no penalty or loss of benefits. However, if you decide to participate, please be informed that there will not be any personal financial gain or academic benefits. You are free to withdraw from this study at any time without stating a reason. If you decide to withdraw after the photos and recordings have been taken, your photos and CCTV footage will immediately be deleted and there will be no penalty.

RISKS AND BENEFITS

There are no risks or side effects involved in this study. You will not benefit directly. This study is of high social and forensic value. Its results can increase the successful prosecution of criminals and identification of missing persons. Which in turn can help solve more police cases and decrease crime on a wider scale.

DATABASE DEVELOPMENT

In the time of taking your photos and recordings, we would also like to create a database of these images. This database will be only used for further research into the topic. It will be assigned a curator from the School of Anatomical Sciences, and will be under the careful watch of the Collections Committee of the School and the curator. No publication or sharing of any of your information will occur at any point throughout the existence of the database, unless the researcher working on the study asks you for further permission, through a re-consent form. You may also decide to participate in the study, but to be excluded from the database if you so desire, at no penalty.

CONFIDENTIALITY

All the information gained during the course of this study is strictly confidential. Your name and contact details will remain confidential with only the photographer and then the primary supervisor (DR N Briers) having access. Your photographs and CCTV footage will be awarded a number. In effect, your personal details will not be linked to any of your photographic material. We require contact details in case we find it necessary to use your photograph in scientific journals or presentations. In this case, you will need to go through a re-consent process, signing an additional form consenting to publication of your photograph. Your photos will only be published with your permission and you may refuse publication

without stating a reason and without any consequences or penalties. No information that could identify you would be used in the thesis or publication, such as name and contact details. The photographer assisting with the study will see your face and details but will not be further involved in the study. A research assistant that will put the photos together into face pools and code the image names, will see your face but will not receive any of your personal information. Two analysts related to the university, including me, will see your photograph as part of the face pools, but will not be given any information about you. I have purposefully excluded myself and the other analyst from taking your photos or information to prevent bias on my later recognition of your face. However I have chosen the geographical and social area of main campus in order to limit accidental recognition by myself or the other analyst. The results of this study will be reported in a thesis and made available to my supervisor and other lecturers in the Faculty of Health Sciences at the University, uploaded to the electronic thesis and dissertation database and may be published in conference presentations and journals.

If you have any questions, please feel free to contact me on 082-457-0843 or my Supervisor, Dr Nanette Briers, on 082-847-4459. To report any complaints or problems, please contact the Chair of the Human Research Ethics Committee, Professor Cleaton-Jones on 011-717-2301 or the Research Administrator, Ms Zanele Ndlovu on 011-717-2700. If you would like, a summary of the research will be made available to you. Thank you for your consideration to participate in this study.

CONSENT FORM – FACIAL PHOTOGRAPHS AND CCTV RECORDINGS

I hereby confirm that I have read and understand the scope of this study and that the investigator, N. Bacci, has explained the nature, conduct, risks and benefits of this study.

I _____ (participant's name) consent to participate in the study "Reliability and admissibility to court of forensic facial identification from CCTV recordings" and consent to have my facial photographs taken, and to be recorded on a CCTV system accessible to the project investigators.

I understand that:

Participation in this study is entirely voluntary.

I can withdraw from this study at any time.

There are no risks or benefits anticipated.

Photographs and recordings will remain strictly confidential and anonymous.

Photographs and recordings will only be published with my permission and consent.

Participant's name _____ Date _____

Participant's signature _____ Contact details _____

Alternative contact details _____

Investigator's name _____ Date _____

Investigator's signature _____

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG
FACULTY OF HEALTH SCIENCES
Phillip V Tobias Health Sciences Building, 29 Princess of Wales Terrance, Parktown 2193; Johannesburg; South Africa T +27 11 717 2000

Conditions of Use Agreement for the Wits Face Database (WFD)

1. Preamble

1.1. The terms and conditions contained in this document govern the right to access and use the Wits Face Database ("WFD") belonging to the University of the Witwatersrand, Johannesburg ("University").

1.2. The participants who have graciously allowed their facial photographs and/or recordings to be captured ("Participants") have done so on the understanding that the WFD will only be used for purposes of non-commercial research in accordance with best academic practice and ethical regulations.

1.3. By accessing, browsing or using the WFD/part thereof, the User agrees and accepts that the terms and conditions contained herein are binding and undertakes to comply with the stipulated terms and conditions.

1.4. A detailed description of the WFD can be found at:
http://doi.org/10.17605/OSF.IO/WMA4C
To see a descriptor and access a sample of the WFD please use the following link: http://wiredspace.wits.ac.za/handle/10539/29924 .

2. Definitions and Interpretations

In this document, unless the context clearly indicates otherwise the following terms are used:

2.1. **"Agreement"** means the terms and conditions of use contained in this document and any annexures and/or schedules hereto.

2.2. **"Applicant"** means anyone who has submitted a formal application for access and to make use of the WFD/part thereof, for non-commercial research purposes;

2.3. **"Collaborator"** means any member of a research group or team authorised to access and to use the WFD/part thereof, for research purposes;

2.4. **"Editor"** means the Editor of the WFD, Nicholas Bacci, who was the principal investigator and creator of the WFD at the University and who is involved in the processes of reviewing applications for access to the WFD;

2.5. "**HoS**" means the Head of the University's School of Anatomical Sciences;

2.6. **"Institutional Review Board"** means the University's School of Anatomical Sciences Collections Committee which acts as a Research Ethics Committee for the purpose of

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG
FACULTY OF HEALTH SCIENCES
Phillip V Tobias Health Sciences Building, 29 Princess of Wales Terrance, Parktown 2193; Johannesburg; South Africa T +27 11 717 2000

the WFD (hereinafter referred to as the "IRB" and "REC" respectively) regulating access to the WFD;

2.7. **"Participants"** refers to the participants who have provided written consent for their facial photographs and/or recordings to be captured for inclusion in the WFD;

2.8. **"Personal Information"** in terms of the Protection of Personal Information Act 4 of 2013 (available at the following URL: https://www.gov.za/documents/protection-personal-information-act), is data that can be used to identify a person. It is defined as "information relating to an identifiable, living, natural person, and where it is applicable, an identifiable, existing juristic person." and includes any data, images, and video recordings of Participants;

2.9. **"School"** means the School of Anatomical Sciences, University of the Witwatersrand, Johannesburg;

2.10. **"Third party**" means any individual, research group, organisation or institution other than the approved User/s;

2.11. **"University"** means the University of the Witwatersrand, Johannesburg;

2.12. **"User"** means a person to who has been granted access to the WFD and who agreed to abide by the terms and conditions of this Agreement and any additional terms and conditions imposed by the IRB/REC**;**

2.13. **"WFD"** refers to the Wits Face Database published under the following DOI: http://doi.org/10.17605/OSF.IO/WMA4C

3. Terms and Conditions of Access and Use

3.1 The WFD is an open access resource for use in strictly non-commercial research. Users may not carry out any work on the WFD unless authorised in writing to do so by the IRB/REC.

3.2 Users and/or Third parties wishing to obtain access to the WFD for research purposes must be affiliated to an academic institution for a minimum period of 6 (six) months prior to the application date. This affiliation, whether as a student or staff member of the institution, may require verification if deemed necessary by the IRB/REC.

3.3 Users and/or Third parties must use their institutional contact details (i.e. email address) for any communications regarding the WFD. These contact details must be verifiable as belonging or being affiliated to the institution.

3.4 Applicants wishing to make use of the WFD must submit their request in writing to the Editor and the IRB/REC. The application must include:

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG
FACULTY OF HEALTH SCIENCES
Phillip V Tobias Health Sciences Building, 29 Princess of Wales Terrance, Parktown 2193; Johannesburg; South Africa T +27 11 717 2000

3.4.1 a succinct research protocol, outlining the aim of the study, the procedures and processes to be employed in collecting data, and specifying the type and quantity of images required as well as how the results will be disseminated and how long data generated will be preserved.

3.4.2 a completed application form as well as a signed copy of this Agreement with name, surname and detailed affiliations of all persons applying for use.

3.5 External Applicants must submit proof of ethical clearance approval from their own institution.

3.6 If the research is for the purposes of a post-graduate degree, evidence of approval of the project proposal from the student's institution must be submitted.

3.7 The facial images and/or recordings of the Participants may not under any circumstances whatsoever be published in any form (e.g. published journal article, news article, thesis/dissertation, conference podium or poster presentation etc.).

3.8 The IRB/REC and Editor will review the application and determine the scientific and ethical merit of the intended research project in order to recommend granting the School's blanket ethics waiver to the HoS. The HoS is the legal custodian of all collections housed within the School, including the WFD, and is authorised to grant the ethics waiver.

3.9 An application can be deemed approved with no changes or approved subject to minor adjustments or requiring re-submission with major edits or rejected if it is in direct violation of any of the terms and conditions stipulated in this document

3.10 Successful Applicants will be informed of the status of the application and will be provided with an ethics template for completion. The ethics template accompanied by a letter from the HoS indicating that he/she has approved the ethics waiver for the study, will be sent to the chair of the Human Research Ethics Committee (Medical) bi-annually for each project approved within the School that makes use of a collection, including the WFD. Once the Applicant has submitted the ethics template, they will be provided directions on how to access the component of the WFD relevant to their study.

4. Copyright and Intellectual Property

4.1 All Intellectual Property rights in the WFD are owned by the University.

4.2 Subject to compliance with the terms and conditions of this Agreement, Users are, for the period of access approved by the IRB/REC hereby granted a temporary, non-

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG
FACULTY OF HEALTH SCIENCES
Phillip V Tobias Health Sciences Building, 29 Princess of Wales Terrance, Parktown 2193; Johannesburg; South Africa T +27 11 717 2000

exclusive, revocable, non-transferable licence to access and to use the WFD/part thereof solely for non-commercial research purposes.

4.3 Such license as may be granted in terms of this Agreement may not be sub-licenced, assigned, pledged, encumbered or otherwise transferred by the User, voluntarily or involuntarily, by operation of law or otherwise, without University's prior written consent, which consent may be withheld in University's sole discretion, and any attempt to do so in violation of this Agreement will be without legal effect and void in terms of this Agreement.

4.4 The data on the WFD is provided "as is", and the University assumes no responsibility for errors or omissions. The User assumes the entire risk associated with its use of these data. The University shall not be held liable for any use or misuse of the data described and/or contained herein. The User bears all responsibility in determining whether these data is fit for the User's intended use.

4.5 For the avoidance of doubt, it is recorded that the WFD/part thereof:

4.5.1 may not be reproduced and/or publicized in any form or nature;

4.5.2 the images and/or recordings and the associated metadata, included in the WFD must not be reproduced without prior written consent of the University and the IRB/REC, which consent may be withheld in University's sole discretion;

4.5.3 the distribution of any constituents of the WFD to any third party is prohibited;

4.5.4 the distribution of any constituents of the WFD for commercial purposes is prohibited; and

4.5.5 as far as possible and where appropriate, projects including the WFD must be carried out in collaboration with the Editor and/or co-developers of the WFD.

4.6 Any screenshots, photographs, reproductions or any other information (hereafter referred to as "Data") obtained from the original facial images and/or recordings belongs to the University. This includes, but is not limited to, data generated through the study of the WFD/part thereof. Data may not be made available on any Internet website, ftp-site, or any other similar public access point.

4.7 Users must comply with all applicable legislation including but not limited to the provisions of the Data Protection Legislation governing the collection, use and processing of Personal Information as defined in Protection of Personal Information Act 4 of 2013 (POPI).

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG
FACULTY OF HEALTH SCIENCES
Phillip V Tobias Health Sciences Building, 29 Princess of Wales Terrance, Parktown 2193; Johannesburg; South Africa T +27 11 717 2000

4.8 Users must not process Personal Information/data obtained from the study of any part of/the whole WFD for any purpose other than in terms of the approved research project, without the express written permission of the IRB/REC. The distribution of data to any Third Party by Users is not permitted. Any Third Party requesting access to the data obtained through a study of the WFD, by a previous User must submit a written request to said User as well as the IRB/REC. Each request will be addressed in an *ad hoc* manner by the Editor and the IRB/REC.

4.9 On completion of the research work, the User undertakes to ensure that copies of all Data and research outputs or part thereof whether in his/her possession or not, is immediately handed over to the Editor.

4.10 Users may not keep copies of the raw data/WFD, delete remove or in any way interfere with the University's Intellectual Property without permission of the IRB/REC. The School reserves the right to use any of these Data in future research and teaching projects. Users will be allowed a moratorium period of maximum 2 (two) years to complete the research goals (i.e. publication, dissertation etc.). After which the Data will be made available for use for competing research projects or teaching. If the User is not able to complete the project within this time period, then the User must submit a progress report stipulating all limiting challenges together with a completion plan.

4.11 This WFD may not in any shape or form be incorporated into any semi- or fully automated facial recognition/detection system. It may, however, be used to test the efficacy of non-commercial systems for face recognition and face detection. All Participants included in this database must be referred to by the commonly accepted terminology and they may not be discussed in any derogatory way.

4.12 Under no circumstances may any Data not strictly belonging to consenting Participants of the study be used for research of any kind. Any passer-by or unintentionally recorded individual's facial or other data may not be used for any purpose, including research. No image or video data may be published in any format or any platform if originating from the WFD, regardless of modifications.

4.13 If a scientific publication results through unauthorised actions of the User and/or Third Party for which approval was not granted in writing by the IRB/REC, the relevant journal/publisher will be notified of this and the journal/publisher will be requested to remove the publication from their issues.

5. Acknowledgement

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG
FACULTY OF HEALTH SCIENCES
Phillip V Tobias Health Sciences Building, 29 Princess of Wales Terrance, Parktown 2193; Johannesburg; South Africa T +27 11 717 2000

5.1 Any publication, book or dissertation / thesis arising from the use of data through the WFD must be sent to the Editor for curation in the School's archive, as close as possible to publication. In all of these publications, clear reference must be made to the DOI of the WFD and the publication describing its composition.

5.2 The funders, contributors and individuals that aided in creating the database both as Participants and volunteers should be acknowledged in all publications, books and theses. The following credit line of acknowledgement must be included:

"The authors would like to thank the School of Anatomical Sciences of the University of the Witwatersrand, Johannesburg, the Wits Library, Protection Services and ICT Services as well as the database participants and the individuals (Gideon LeRoux, Nina Lewin, Mark Allen) who assisted in the development and curation of the database. The WFD was developed with support from the South African National Research Foundation (Grant Nos.: 11858; CSUR160425163022 (UID:106031)) and JJJ Smieszeck Fellowship. Any opinions, findings and conclusions or recommendations expressed in this study are those of the authors and therefore the NRF and University of the Witwatersrand, Johannesburg do not accept any liability in regard thereto."

6. Breach

6.1 Failure to comply with the terms and condition of this Agreement, will result in the User's/Third Party's and their institution's licence granted in terms of clause 4.2 being immediately revoked. Future access to the WFD and other collections or databases that fall under the management of the School will be denied. In addition, the academic institution of any Users in breach of this document will be notified of the User's unethical and illegal conduct.

6.2 The University will be entitled, without prejudice to any other rights it may have in terms of this Agreement or at law, to terminate this Agreement and to claim any damages it may have suffered as a result of the breach.

6.3 Notwithstanding anything contained herein, the cancellation of this Agreement shall not affect the University's rights that had accrued at the date of termination.

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG
FACULTY OF HEALTH SCIENCES
Phillip V Tobias Health Sciences Building, 29 Princess of Wales Terrance, Parktown 2193; Johannesburg; South Africa T +27 11 717 2000

7. Indemnity

Users hereby indemnify and hold the University harmless from any claim which may arise in terms of this Agreement. Users/Third Parties will not under any circumstances hold the University liable for any indirect, contingent or consequential loss sustained or incurred by them howsoever arising and of whatsoever nature, including but not limited to loss of profits, whether or not the Users/Third Parties contemplated such losses or damages at any time.

8. Domicilium Citandi et Executandi

The Parties select as their respective domicilia citandi et executandi for the purposes of legal proceedings and for the purposes of giving or sending notices provided for or necessary in terms of this Agreement, the following addresses:

8.1 **The University:**

8.1.1 Operational Notices

Attention: Collections Committee of the School of Anatomical Sciences

Physical address: 7 York Road

2nd Floor, Wits Medical School

Parktown,

Johannesburg

2193

8.1.2 Legal Notices

The Director: Legal Services

Physical address: 1 Jan Smuts Avenue

5th Floor Solomon Mahlangu House

Braamfontein

Johannesburg

2001

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG
FACULTY OF HEALTH SCIENCES
Phillip V Tobias Health Sciences Building, 29 Princess of Wales Terrance, Parktown 2193; Johannesburg; South Africa T +27 11 717 2000

8.2 **The User:**

8.2.1 Operational Notices

Attention:

Physical address:

8.2.2 Legal Notices

Attention:

Physical address:

I acknowledge that I have read and understood the terms and conditions stipulated in this Agreement and I undertake to abide by the terms and conditions contained herein.

Signed at_____ on this _____ day of _____20___ .

_____

Principal Investigator

Full name: _____

Affiliation: _____

Signed at_____ on this _____ day of _____20___ .

_____

Collaborator #1

Full name: _____

Affiliation: _____

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG
FACULTY OF HEALTH SCIENCES
Phillip V Tobias Health Sciences Building, 29 Princess of Wales Terrance, Parktown 2193; Johannesburg; South Africa T +27 11 717 2000

Signed at_____ on this _____ day of _____20__ .

_____

Collaborator #2

Full name: _____

Affiliation: _____

Signed at_____ on this _____ day of _____20__ .

_____

Collaborator #3

Full name: _____

Affiliation: _____

Signed at_____ on this _____ day of _____20__ .

_____

Collaborator #4

Full name: _____

Affiliation: _____

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG
FACULTY OF HEALTH SCIENCES
Phillip V Tobias Health Sciences Building, 29 Princess of Wales Terrance, Parktown 2193; Johannesburg; South Africa T +27 11 717 2000

Signed at_____ on this _____ day of _____20__ .

_____

Collaborator #5

Full name: _____

Affiliation: _____

Signed at_____ on this _____ day of _____20__ .

_____

First Witness

Full name: _____

Affiliation: _____

Signed at_____ on this _____ day of _____20__ .

_____

Second Witness

Full name: _____

Affiliation: _____

Appendix D - Supplementary Material Manuscript 2

Supplementary Information: equation list drawn from the Caret R package [1, 2]

Equation 1:

$$Sensitivity\ (Recall) = \frac{TP}{(TP + FN)}$$

Equation 2:

$$Specificity = \frac{TN}{(FP + TN)}$$

Equation 3:

$$Prevalence = \frac{(TP + FN)}{(TP + FP + FN + TN)}$$

Equation 4:

$$PPV = \frac{(Sensitivity \times Prevalence)}{\{(Sensitivity \times Prevalence) + [(1 - Specificity) \times (1 - Prevalence)]\}}$$

Equation 5:

$$NPV = \frac{[Specificity \times (1 - Prevalence)]}{\{[1 - Sensitivity) \times Prevalence] + [Specificity \times (1 - Prevalence)]\}}$$

Equation 6:

$$Detection\ Rate = \frac{TP}{(TP + FP + FN + TN)}$$

Equation 7:

$$Detection\ Prevalence = \frac{(TP + FP)}{(TP + FP + FN + TN)}$$

Equation 8:

$$Balanced\ Accuracy = \frac{(Sensitivity + Specificity)}{2}$$

Equation 9:

$$Precision = \frac{TP}{(TP + FP)}$$

Equation 10 (where β = 1):

$$F_1\text{-Score} = \frac{(1 + \beta^2) \times Precision\ \times Sensitivity}{(\beta^2 \times Precision) + Sensitivity}$$

1. Kuhn M (2020) Caret: classification and regression training R package
2. Kuhn M (2008) Building predictive models in R using the caret package. J Stat Softw 28:1–26. https://doi.org/10.18637/jss.v028.i05

Supplementary Information – Summary of individual outcomes by face pool in table format

Table SI1: Summary of photo-to-photo comparison outcomes by face pool

	Lenient Iteration					Strict Iteration				
Face Pool Number	TP	FP	TN	FN	IA	TP	FP	TN	FN	IA
1	1	0	9	0	0	1	0	9	0	0
2	1	0	9	0	0	1	0	9	0	0
3	1	0	9	0	0	1	0	9	0	0
4	1	0	9	0	0	1	0	9	0	0
5	1	0	9	0	0	1	0	9	0	0
6	1	0	9	0	0	1	0	9	0	0
7	1	0	9	0	0	1	0	9	0	0
8	1	0	9	0	0	1	0	9	0	0
9	1	0	9	0	0	1	0	9	0	0
10	1	0	9	0	0	1	0	9	0	0
11	1	0	9	0	0	1	0	9	0	0
12	1	0	9	0	0	1	0	9	0	0
13	1	0	9	0	0	1	0	9	0	0
14	1	1	8	0	0	1	0	9	0	0
15	1	0	9	0	0	1	0	9	0	0
16	1	0	9	0	0	1	0	9	0	0
17	1	0	9	0	0	1	0	9	0	0
18	1	0	9	0	0	1	0	9	0	0
19	1	0	9	0	0	1	0	9	0	0
20	1	0	9	0	0	1	0	9	0	0
21	1	0	9	0	0	1	0	9	0	0
22	1	0	9	0	0	1	0	9	0	0
23	1	0	9	0	0	1	0	9	0	0
24	1	0	9	0	0	1	0	9	0	0
25	1	0	9	0	0	1	0	9	0	0
26	0	0	10	0	0	0	0	10	0	0
27	1	0	9	0	0	1	0	9	0	0
28	1	0	9	0	0	1	0	9	0	0
29	1	0	9	0	0	1	0	9	0	0
30	1	0	9	0	0	1	0	9	0	0
31	1	0	9	0	0	1	0	9	0	0
32	1	0	9	0	0	1	0	9	0	0
33	1	0	9	0	0	1	0	9	0	0
34	1	0	9	0	0	1	0	9	0	0
35	1	0	9	0	0	1	0	9	0	0
36	1	0	9	0	0	1	0	9	0	0
37	1	0	9	0	0	1	0	9	0	0
38	0	0	10	0	0	0	0	10	0	0
39	0	0	9	1	0	0	0	9	1	0
40	1	0	9	0	0	1	0	9	0	0
41	1	0	9	0	0	1	0	9	0	0

42	1	0	9	0	0	1	0	9	0	0
43	1	0	9	0	0	1	0	9	0	0
44	0	0	10	0	0	0	0	10	0	0
45	1	0	9	0	0	1	0	9	0	0
46	1	0	9	0	0	1	0	9	0	0
47	1	0	9	0	0	1	0	9	0	0
48	0	0	10	0	0	0	0	10	0	0
49	1	0	9	0	0	1	0	9	0	0
50	1	0	9	0	0	1	0	9	0	0
51	1	0	9	0	0	1	0	9	0	0
52	1	0	9	0	0	1	0	9	0	0
53	1	0	9	0	0	1	0	9	0	0
54	1	0	9	0	0	1	0	9	0	0
55	1	0	9	0	0	1	0	9	0	0
56	1	0	9	0	0	1	0	9	0	0
57	1	0	9	0	0	1	0	9	0	0
58	1	0	9	0	0	1	0	9	0	0
59	1	0	9	0	0	1	0	9	0	0
60	1	0	9	0	0	1	0	9	0	0
61	1	0	9	0	0	1	0	9	0	0
62	1	0	9	0	0	1	0	9	0	0
63	1	0	9	0	0	1	0	9	0	0
64	1	0	9	0	0	1	0	9	0	0
65	0	0	10	0	0	0	0	10	0	0
66	1	0	9	0	0	1	0	9	0	0
67	1	0	9	0	0	1	0	9	0	0
68	1	0	9	0	0	1	0	9	0	0
69	1	0	9	0	0	1	0	9	0	0
70	1	0	9	0	0	1	0	9	0	0
71	1	0	9	0	0	1	0	9	0	0
72	1	0	9	0	0	1	0	9	0	0
73	1	0	9	0	0	1	0	9	0	0
74	1	0	9	0	0	1	0	9	0	0
75	1	0	9	0	0	1	0	9	0	0

Table SI2: Summary of CCTV to photo comparison outcomes by face pool

	Lenient Iteration					Strict Iteration				
Face Pool Number	TP	FP	TN	FN	IA	TP	FP	TN	FN	IA
1	1	0	9	0	0	1	0	9	0	0
2	1	0	9	0	0	1	0	9	0	0
3	1	0	9	0	0	1	0	9	0	0
4	1	0	9	0	0	1	0	9	0	0
5	1	0	9	0	0	1	0	9	0	0
6	1	0	9	0	0	1	0	9	0	0
7	1	0	9	0	0	1	0	9	0	0
8	1	0	9	0	0	1	0	9	0	0
9	1	0	9	0	0	1	0	9	0	0
10	0	0	10	0	0	0	0	10	0	0
11	1	0	9	0	0	1	0	9	0	0
12	1	0	9	0	0	1	0	9	0	0
13	1	0	9	0	0	1	0	9	0	0
14	1	0	7	2	0	0	1	8	1	0
15	0	0	9	1	0	0	0	9	1	0
16	1	0	8	1	0	1	0	9	0	0
17	1	0	9	0	0	1	0	9	0	0
18	1	0	8	1	0	1	0	9	0	0
19	1	0	9	0	0	1	0	9	0	0
20	1	0	8	1	0	0	1	9	0	0
21	1	0	8	1	0	1	0	9	0	0
22	1	0	9	0	0	1	0	9	0	0
23	1	0	9	0	0	1	0	9	0	0
24	0	1	9	0	0	0	1	9	0	0
25	1	0	9	0	0	1	0	9	0	0
26	1	0	9	0	0	1	0	9	0	0
27	0	0	7	3	0	0	0	9	1	0
28	1	0	9	0	0	1	0	9	0	0
29	1	0	9	0	0	1	0	9	0	0
30	0	1	8	1	0	0	1	9	0	0
31	1	0	9	0	0	1	0	9	0	0
32	1	0	9	0	0	0	1	9	0	0
33	1	0	9	0	0	1	0	9	0	0
34	1	0	9	0	0	1	0	9	0	0
35	1	0	9	0	0	1	0	9	0	0
36	1	0	8	1	0	1	0	9	0	0
37	1	0	9	0	0	1	0	9	0	0
38	1	0	9	0	0	1	0	9	0	0
39	1	0	9	0	0	1	0	9	0	0
40	1	0	8	1	0	1	0	9	0	0
41	1	0	8	1	0	1	0	9	0	0
42	1	0	9	0	0	1	0	9	0	0

43	1	0	9	0	0	1	0	9	0	0
44	1	0	8	1	0	1	0	9	0	0
45	1	0	9	0	0	1	0	9	0	0
46	1	0	8	1	0	1	0	9	0	0
47	1	0	9	0	0	1	0	9	0	0
48	1	0	9	0	0	1	0	9	0	0
49	1	0	9	0	0	1	0	9	0	0
50	1	0	9	0	0	1	0	9	0	0
51	1	0	9	0	0	1	0	9	0	0
52	1	0	9	0	0	1	0	9	0	0
53	1	0	9	0	0	1	0	9	0	0
54	1	0	9	0	0	1	0	9	0	0
55	1	0	9	0	0	1	0	9	0	0
56	1	0	9	0	0	1	0	9	0	0
57	1	0	9	0	0	1	0	9	0	0
58	1	0	9	0	0	1	0	9	0	0
59	1	0	9	0	0	1	0	9	0	0
60	1	0	8	1	0	1	0	9	0	0
61	1	0	9	0	0	1	0	9	0	0
62	1	0	9	0	0	1	0	9	0	0
63	1	0	8	1	0	1	0	9	0	0
64	1	0	8	1	0	1	0	9	0	0
65	0	0	10	0	0	0	0	10	0	0
66	1	0	9	0	0	1	0	9	0	0
67	1	0	8	1	0	1	0	9	0	0
68	0	1	9	0	0	0	1	9	0	0
69	1	0	8	1	0	1	0	9	0	0
70	1	0	8	1	0	1	0	9	0	0
71	1	0	9	0	0	1	0	9	0	0
72	1	0	9	0	0	1	0	9	0	0
73	0	0	9	1	0	0	0	10	0	0
74	1	0	9	0	0	0	1	9	0	0
75	0	1	8	1	0	0	1	8	1	0
76	1	0	7	2	0	0	1	9	0	0
77	0	0	0	0	10	0	0	0	0	10
78	1	0	8	1	0	0	1	9	0	0
79	1	0	9	0	0	1	0	9	0	0
80	0	0	1	0	9	0	0	1	0	9
81	0	1	7	2	0	0	1	9	0	0
82	0	1	8	1	0	0	1	9	0	0
83	1	0	9	0	0	0	1	9	0	0
84	1	0	7	2	0	1	0	9	0	0
85	0	1	8	1	0	0	1	9	0	0
86	1	0	7	2	0	1	0	9	0	0
87	0	1	7	1	1	0	1	8	0	1

88	1	0	7	2	0	0	1	8	1	0
89	1	0	7	2	0	1	0	9	0	0
90	1	0	9	0	0	1	0	9	0	0
91	1	0	9	0	0	1	0	9	0	0
92	1	0	8	1	0	1	0	9	0	0
93	0	1	6	3	0	0	1	9	0	0
94	1	0	9	0	0	1	0	9	0	0
95	0	0	9	0	1	0	0	9	0	1
96	0	1	8	1	0	0	1	9	0	0
97	1	0	8	1	0	1	0	9	0	0
98	1	0	9	0	0	0	1	9	0	0
99	1	0	9	0	0	0	1	9	0	0
100	0	1	4	0	5	0	1	4	0	5

Appendix E - Supplementary Material Manuscript 3

Supplementary Information: equation list drawn from the Caret R package [1, 2]

Equation 1:

$$Sensitivity\ (Recall) = \frac{TP}{(TP + FN)}$$

Equation 2:

$$Specificity = \frac{TN}{(FP + TN)}$$

Equation 3:

$$Prevalence = \frac{(TP + FN)}{(TP + FP + FN + TN)}$$

Equation 4:

$$PPV = \frac{(Sensitivity \times Prevalence)}{\{(Sensitivity \times Prevalence) + [(1 - Specificity) \times (1 - Prevalence)]\}}$$

Equation 5:

$$NPV = \frac{[Specificity \times (1 - Prevalence)]}{\{[1 - Sensitivity) \times Prevalence] + [Specificity \times (1 - Prevalence)]\}}$$

Equation 6:

$$Detection\ Rate = \frac{TP}{(TP + FP + FN + TN)}$$

Equation 7:

$$Detection\ Prevalence = \frac{(TP + FP)}{(TP + FP + FN + TN)}$$

Equation 8:

$$Balanced\ Accuracy = \frac{(Sensitivity + Specificity)}{2}$$

Equation 9:

$$Precision = \frac{TP}{(TP + FP)}$$

Equation 10 (where β = 1):

$$F_1\text{-Score} = \frac{(1 + \beta^2) \times Precision\ \times Sensitivity}{(\beta^2 \times Precision) + Sensitivity}$$

1. Kuhn M (2020) Caret: classification and regression training R package
2. Kuhn M (2008) Building predictive models in R using the caret package. J Stat Softw 28:1–26. https://doi.org/10.18637/jss.v028.i05

Appendix F - Supplementary Material Manuscript 4

Supplementary Information: equation list drawn from the Caret R package (1,2)

Equation 1:

$$Sensitivity\ (Recall) = \frac{TP}{(TP + FN)}$$

Equation 2:

$$Specificity = \frac{TN}{(FP + TN)}$$

Equation 3:

$$Prevalence = \frac{(TP + FN)}{(TP + FP + FN + TN)}$$

Equation 4:

$$PPV = \frac{(Sensitivity \times Prevalence)}{\{(Sensitivity \times Prevalence) + [(1 - Specificity) \times (1 - Prevalence)]\}}$$

Equation 5:

$$NPV = \frac{[Specificity \times (1 - Prevalence)]}{\{[1 - Sensitivity) \times Prevalence] + [Specificity \times (1 - Prevalence)]\}}$$

Equation 6:

$$Detection\ Rate = \frac{TP}{(TP + FP + FN + TN)}$$

Equation 7:

$$Detection\ Prevalence = \frac{(TP + FP)}{(TP + FP + FN + TN)}$$

Equation 8:

$$Balanced\ Accuracy = \frac{(Sensitivity + Specificity)}{2}$$

Equation 9 (where β = 1):

$$F_1\text{-Score} = \frac{(1 + \beta^2) \times Precision\ \times Sensitivity}{(\beta^2 \times Precision) + Sensitivity}$$

1. Kuhn M. Caret: classification and regression training R package. 2020.
2. Kuhn M. Building predictive models in R using the caret package. J Stat Softw 2008;28(5):1–26. https://doi.org/10.18637/jss.v028.i05.

Appendix F - Publication Licences & Copyright

F1000 Research Licenses (accessed at: https://f1000research.com/about)

The licenses that apply to articles and other research outputs, data and peer review reports. F1000Research articles are usually published under a CC BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited, and leaves the copyright of the article with the current copyright holder (usually the author or their institution). Additional waivers are used by some institutions, for example governmental employees, where appropriate. As the specific version of the CC BY license applied to articles may change due to periodic updates, the copyright information for each article is shown below the abstract. Most posters, slides and documents are also published under a CC BY license, but other CC licenses may apply, as indicated for each research output's published page under the image. Data associated with F1000Research articles (not posters, slides and documents unless specifically stated) are made available, where possible, under the terms of a Creative Commons Public Domain Dedication (CC0 license). This facilitates and encourages data re-use and helps prevent the problems of attribution stacking when combining multiple datasets each authored by multiple authors that use multiple different licenses. Peer review reports that are published with a given article are available under the CC BY license.

Corresponding author: Nicholas Bacci

Competing interests: No competing interests were disclosed.

Grant information: The research of N. Bacci (Grant No.: 11858) and N. Briers (Grant No.: CSUR160425163022 (UID:106031)) is sponsored by the National Research Foundation of South Africa. Any opinions, findings and conclusions or recommendations expressed in this study are those of the authors and therefore the NRF does not accept any liability in regard thereto. N. Bacci was also partially funded by the J.J.J. Smieszek Fellowship, School of Anatomical Sciences, University of the Witwatersrand.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 Bacci N *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

How to cite: Bacci N, Davimes J, Steyn M and Briers N. Development of the Wits Face Database: an African database of high-resolution facial photographs and multimodal closed-circuit television (CCTV) recordings [version 1; peer review: 1 approved]. *F1000Research* 2021, 10:131 (https://doi.org/10.12688/f1000research.50887.1)

First published: 19 Feb 2021, 10:131 (https://doi.org/10.12688/f1000research.50887.1)
Latest published: 19 Feb 2021, 10:131 (https://doi.org/10.12688/f1000research.50887.1)

Springer Nature - International Journal of Legal Medicine License

Licence to Publish

SPRINGER NATURE

Journal Name:	International Journal of Legal Medicine	(the 'Journal')
Manuscript Number:	IJLM-D-20-00713R1	
Proposed Title of Article:	Validation of forensic facial comparison by morphological analysis in photographic and CCTV samples	
Author(s) [Please list all named Authors]:	Nicholas Bacci, Tobias M. R. Houlton, Nanette Briers, Maryna Steyn	(the 'Author')
Corresponding Author Name:	Nicholas Bacci	

1 Publication

Springer-Verlag GmbH, DE (the 'Licensee') will consider publishing this article, including any supplementary information and graphic elements therein (e.g. illustrations, charts, moving images) (the 'Article').
Headings are for convenience only.

2 Grant of Rights

In consideration of the Licensee evaluating the Article for publication, the Author grants the Licensee the exclusive (except as set out in clauses 3, 4 and 5a) iv) and sub-licensable right, unlimited in time and territory, to copy-edit, reproduce, publish, distribute, transmit, make available and store the Article, including abstracts thereof, in all forms of media of expression now known or developed in the future, including pre- and reprints, translations, photographic reproductions and extensions.

Furthermore, to enable additional publishing services, such as promotion of the Article, the Author grants the Licensee the right to use the Article (including the use of any graphic elements on a stand-alone basis) in whole or in part in electronic form, such as for display in databases or data networks (e.g. the Internet), or for print or download to stationary or portable devices. This includes interactive and multimedia use as well as posting the Article in full or in part or its abstract on social media, and the right to alter the Article to the extent necessary for such use. The Licensee may also let third parties share the Article in full or in part or its abstract on social media and may in this context sub-license the Article and its abstract to social media users. Author grants to Licensee the right to re-license Article metadata without restriction (including but not limited to author name, title, abstract, citation, references, keywords and any additional information as determined by Licensee).

If the Article is rejected by the Licensee and not published, all rights under this agreement shall revert to the Author.

3 Self Archiving

Author is permitted to self-archive a preprint and the accepted manuscript version of their Article.

a) A preprint is the version of the Article before peer-review has taken place ("Preprint"). Prior to acceptance for publication, Author retains the right to make a Preprint of their Article available on any of the following: their own personal, self-maintained website; a legally compliant Preprint server such as but not limited to arXiv and bioRxiv. Once the Article has been published, the Author should update the acknowledgement and provide a link to the definitive version on the publisher's website: "This is a preprint of an article published in [insert journal title]. The final authenticated version is available online at: https://doi.org/[insert DOI]"

b) The accepted manuscript version, by industry standard called the "Author's Accepted Manuscript" ("AAM") is the version accepted for publication in a journal following peer review but prior to copyediting and typesetting that can be made available under the following conditions:
(i) Author retains the right to make an AAM of the Article available on their own personal, self-maintained website immediately on acceptance, (ii) Author retains the right to make an AAM of the Article available for public release on any of the following 12 months after first publication ("Embargo Period"): their employer's internal website; their institutional and/or funder repositories; AAMs may also be deposited in such repositories immediately on acceptance, provided that they are not made publicly available until after the Embargo Period.

An acknowledgement in the following form should be included, together with a link to the published version on the publisher's website: "This is a post-peer-review, pre-copyedit version of an article published in [insert journal title]. The final authenticated version is available online at: http://dx.doi.org/[insert DOI]".

4 Reuse Rights

Author retains the following non-exclusive rights for the published version provided that, when reproducing the Article or extracts from it, the Author acknowledges and references first publication in the Journal according to current citation standards. In any event the acknowledgement should contain as a minimum, "First published in [Journal name, volume, page number, year] by Springer Nature".

a) to reuse graphic elements created by the Author and contained in the Article, in presentations and other works created by them;

b) the Author and any academic institution where they work at the time may reproduce the Article for the purpose of course teaching (but not for inclusion in course pack material for onward sale by libraries and institutions);

c) to reuse the published version of the Article or any part in a thesis written by the same Author , and to make a copy of that thesis available in a repository of the Author(s)' awarding academic institution, or other repository required by the awarding academic institution. An acknowledgement should be included in the citation: "Reproduced with permission from Springer Nature"; and

d) to reproduce, or to allow a third party to reproduce the Article, in whole or in part, in any other type of work (other than thesis) written by the Author for distribution by a publisher after an embargo period of 12 months.

5 Warranties & Representations

Author warrants and represents that:

a)
- i. the Author is the sole copyright owner or has been authorised by any additional copyright owner(s) to grant the rights defined in clause 2,
- ii. the Article does not infringe any intellectual property rights (including without limitation copyright, database rights or trade mark rights) or other third party rights and no licence from or payments to a third party are required to publish the Article,
- iii. the Article has not been previously published or licensed,
- iv. if the Article contains materials from other sources (e.g. illustrations, tables, text quotations), Author has obtained written permissions to the extent necessary from the copyright holder(s), to license to the Licensee the same rights as set out in clause 2 and has cited any such materials correctly;

b) all of the facts contained in the Article are according to the current body of research true and accurate;

c) nothing in the Article is obscene, defamatory, violates any right of privacy or publicity, infringes any other human, personal or other rights of any person or entity or is otherwise unlawful and that informed consent to publish has been obtained for all research participants;

d) nothing in the Article infringes any duty of confidentiality which Author might owe to anyone else or violates any contract, express or implied, of Author. All of the institutions in which work recorded in the Article was created or carried out have authorised and approved such research and publication; and

e) the signatory who has signed this agreement has full right, power and authority to enter into this agreement on behalf of all of the Authors.

6 Cooperation

a) Author shall cooperate fully with the Licensee in relation to any legal action that might arise from the publication of the Article, and the Author shall give the Licensee access at reasonable times to any relevant accounts, documents and records within the power or control of the Author. Author agrees that the distributing entity is intended to have the benefit of and shall have the right to enforce the terms of this agreement.

b) Author authorises the Licensee to take such steps as it considers necessary at its own expense in the Author's name(s) and on their behalf if the Licensee believes that a third party is infringing or is likely to infringe copyright in the Article including but not limited to initiating legal proceedings.

7 Author List

Changes of authorship, including, but not limited to, changes in the corresponding author or the sequence of authors, are not permitted after acceptance of a manuscript.

8 Corrections

Author agrees that the Licensee may retract the Article or publish a correction or other notice in relation to the Article if the Licensee considers in its reasonable opinion that such actions are appropriate from a legal, editorial or research integrity perspective.

9 Governing Law

This agreement shall be governed by, and shall be construed in accordance with, the laws of the Federal Republic of Germany. The courts of Berlin, Germany shall have the exclusive jurisdiction.

Wiley – Journal of Forensic Sciences License Agreement

Journal of Forensic Sciences

Published by Wiley on behalf of American Academy of Forensic Sciences (the "Owner")

COPYRIGHT TRANSFER AGREEMENT

Date: March 24, 2021

Contributor name: Nicholas Bacci

Contributor address:

Manuscript number: JOFS-21-068.R1

Re: Manuscript entitled Assessing the effect of facial disguises on forensic facial comparison by morphological analysis (the "Contribution")

for publication in Journal of Forensic Sciences (the "Journal")

published by Wiley Periodicals LLC ("Wiley")

Dear Contributor(s):

Thank you for submitting your Contribution for publication. In order to expedite the editing and publishing process and enable the Owner to disseminate your Contribution to the fullest extent, we need to have this Copyright Transfer Agreement executed. If the Contribution is not accepted for publication, or if the Contribution is subsequently rejected, this Agreement shall be null and void.
Publication cannot proceed without a signed copy of this Agreement.

A. COPYRIGHT

1. The Contributor assigns to the Owner, during the full term of copyright and any extensions or renewals, all copyright in and to the Contribution, and all rights therein, including but not limited to the right to publish, republish, transmit, sell, distribute and otherwise use the Contribution in whole or in part in electronic and print editions of the Journal and in derivative works throughout the world, in all languages and in all media of expression now known or later developed, and to license or permit others to do so. For the avoidance of doubt, "Contribution" is defined to only include the article submitted by the Contributor for publication in the Journal (including any embedded rich media) and does not extend to any supporting information submitted with or referred to in the Contribution ("Supporting Information"). To the extent that any Supporting Information is submitted to the Journal, the Owner is granted a perpetual, non-exclusive license to publish, republish, transmit, sell, distribute and otherwise use this Supporting Information in whole or in part in electronic and print editions of the Journal and in derivative works throughout the world, in all languages and in all media of expression now known or later developed, and to license or permit others to do so. If the Contribution was shared as a preprint, the Contributor grants to the Owner exclusivity as to any rights retained by the Contributor in the preprint.

2. Reproduction, posting, transmission or other distribution or use of the final Contribution in whole or in part in any medium by the Contributor as permitted by this Agreement requires a citation to the Journal suitable in form and content as follows: (Title of Article, Contributor, Journal Title and Volume/Issue, Copyright © [year], copyright owner as specified in the Journal, Publisher). Links to the final article on the publisher website are encouraged where appropriate.

B. RETAINED RIGHTS

Notwithstanding the above, the Contributor or, if applicable, the Contributor's employer, retains all proprietary rights other than copyright, such as patent rights, in any process, procedure or article of manufacture described in the Contribution.

C. PERMITTED USES BY CONTRIBUTOR

1. Submitted Version. The Owner licenses back the following rights to the Contributor in the version of the Contribution as originally submitted for publication (the "Submitted Version"):

a. The right to self-archive the Submitted Version on: the Contributor's personal website; a not for profit subject-based preprint server or repository; a Scholarly Collaboration Network (SCN) which has signed up to the STM article sharing principles [http://www.stm-assoc.org/stm-consultations/scn-consultation-2015/] ("Compliant SCNs"); or the Contributor's company/ institutional repository or archive. This right extends to both intranets and the Internet. The Contributor may replace the Submitted Version with the Accepted Version, after any relevant embargo period as set out in paragraph C.2(a) below has elapsed. The Contributor may wish to add a note about acceptance by the Journal and upon publication it is recommended that Contributors add a Digital Object Identifier (DOI) link back to the Final Published Version.

b. The right to transmit, print and share copies of the Submitted Version with colleagues, including via Compliant SCNs, provided that there is no systematic distribution of the Submitted Version, e.g. posting on a listserve, network (including SCNs which have not signed up to the STM sharing principles) or automated delivery.

2. Accepted Version. The Owner licenses back the following rights to the Contributor in the version of the Contribution that has been peer-reviewed and accepted for publication, but not final (the "Accepted Version"):

a. The right to self-archive the Accepted Version on: the Contributor's personal website; the Contributor's company/institutional repository or archive; Compliant SCNs; and not for profit subject-based repositories such as PubMed Central, all subject to an embargo period of 12 months for scientific, technical and medical (STM) journals and 24 months for social science and humanities (SSH) journals following publication of the Final Published Version. There are separate arrangements with certain funding agencies governing reuse of the Accepted Version as set forth at the following website: http://www.wileyauthors.com/funderagreements. The Contributor may not update the Accepted Version or replace it with the Final Published Version. The Accepted Version posted must contain a legend as follows: This is the accepted version of the following article: FULL CITE, which has been published in final form at [Link to final article]. This article may be used for non-commercial purposes in accordance with the Wiley Self-Archiving Policy [http://www.wileyauthors.com/self-archiving].

b. The right to transmit, print and share copies of the Accepted Version with colleagues, including via Compliant SCNs (in private research groups only before the embargo and publicly after), provided that there is no systematic distribution of the Accepted Version, e.g. posting on a listserve, network (including SCNs which have not signed up to the STM sharing principles) or automated delivery.

3. Final Published Version. The Owner hereby licenses back to the Contributor the following rights with respect to the final published version of the Contribution (the "Final Published Version"):

a. Copies for colleagues. The personal right of the Contributor only to send or transmit individual copies of the Final Published Version in any format to colleagues upon their specific request, and to share copies in private sharing groups in Compliant SCNs, provided no fee is charged, and further provided that there is no systematic external or public distribution of the Final Published Version, e.g. posting on a listserve, network or automated delivery.

b. Re-use in other publications. The right to re-use the Final Published Version or parts thereof for any publication authored or edited by the Contributor (excluding journal articles) where such re-used material constitutes less than half of the total material in such publication. In such case, any modifications must be accurately noted.

c. Teaching duties. The right to include the Final Published Version in teaching or training duties at the Contributor's institution/place of employment including in course packs, e-reserves, presentation at professional conferences, in-house training, or distance learning. The Final Published Version may not be used in seminars outside of normal teaching obligations (e.g. commercial seminars). Electronic posting of the Final Published Version in connection with teaching/training at the Contributor's company/institution is permitted subject to the implementation of reasonable access control mechanisms, such as user name and password. Posting the Final Published Version on the open Internet is not permitted.

d. Oral presentations. The right to make oral presentations based on the Final Published Version.

4. Article Abstracts, Figures, Tables, Artwork and Selected Text (up to 250 words).

a. Contributors may re-use unmodified abstracts for any non-commercial purpose. For online uses of the abstracts, the Owner encourages but does not require linking back to the Final Published Version.

b. Contributors may re-use figures, tables, artwork, and selected text up to 250 words from their Contributions, provided the following conditions are met:

(i) Full and accurate credit must be given to the Final Published Version.

(ii) Modifications to the figures and tables must be noted. Otherwise, no changes may be made.

(iii) The re-use may not be made for direct commercial purposes, or for financial consideration to the Contributor.

(iv) Nothing herein will permit dual publication in violation of journal ethical practices.

D. CONTRIBUTIONS OWNED BY EMPLOYER

1. If the Contribution was written by the Contributor in the course of the Contributor's employment as a "work-made-for-hire" in the course of employment, the Contribution is owned by the company/institution which must execute this Agreement (in addition to the Contributor's signature). In such case, the company/institution hereby agrees to the terms of use set forth in paragraph A above and assigns to the Owner, during the full term of copyright, all copyright in and to the Contribution for the full term of copyright throughout the world as specified in paragraph A above.

2. In addition to the rights specified as retained in paragraph B above and the rights granted back to the Contributor pursuant to paragraph C above, the Owner hereby grants back, without charge, to such company/institution, its subsidiaries and divisions, the right to make copies of and distribute the Final Published Version internally in print format

or electronically on the Company's internal network. Copies so used may not be resold or distributed externally. However, the company/institution may include information and text from the Final Published Version as part of an information package included with software or other products offered for sale or license or included in patent applications. Posting of the Final Published Version by the company/institution on a public access website may only be done with written permission, and payment of any applicable fee(s). Also, upon payment of the applicable reprint fee, the company/institution may distribute print copies of the Final Published Version externally.

E. GOVERNMENT CONTRACTS

In the case of a Contribution prepared under U.S. Government contract or grant, the U.S. Government may reproduce, without charge, all or portions of the Contribution and may authorize others to do so, for official U.S. Government purposes only, if the U.S. Government contract or grant so requires. (U.S. Government, U.K. Government, and other government employees: see notes at end.)

F. COPYRIGHT NOTICE

The Contributor and the company/institution agree that any and all copies of the Final Published Version or any part thereof distributed or posted by them in print or electronic format as permitted herein will include the notice of copyright as stipulated in the Journal and a full citation to the Journal.

G. CONTRIBUTOR'S REPRESENTATIONS

The Contributor represents that: (i) the Contributor and all co-Contributors have the full power, authority and capability to enter into this Agreement, to grant the rights and license granted herein and to perform all obligations hereunder; (ii) neither the Contributor nor any co-Contributor has granted exclusive rights to, or transferred their copyright in, any version of the Contribution to any third party; (iii) the Contribution is the Contributor's original work, all individuals identified as Contributors actually contributed to the Contribution, and all individuals who contributed are included; (iv) if the Contribution was prepared jointly, the Contributor has informed the co-Contributors of the terms of this Agreement and has obtained their signed written permission to execute this Agreement on their behalf; (v) the Contribution is submitted only to this Journal and has not been published before, has not been included in another manuscript, and is not currently under consideration or accepted for publication elsewhere; (vi) if excerpts from copyrighted works owned by third parties are included, the Contributor shall obtain written permission from the copyright owners for all uses as set forth in the standard permissions form and the Journal's Author Guidelines, and show credit to the sources in the Contribution; (vii) the Contribution and any submitted Supporting Information contain no libelous or unlawful statements, do not infringe upon the rights (including without limitation the copyright, patent or trademark rights) or the privacy of others, do not breach any confidentiality obligation, do not violate a contract or any law, do not contain material or instructions that might cause harm or injury, and only utilize data that has been obtained in accordance with applicable legal requirements and Journal policies; and (viii) there are no conflicts of interest relating to the Contribution, except as disclosed. Accordingly, the Contributor represents that the following information shall be clearly identified on the title page of the Contribution: (1) all financial and material support for the research and work; (2) any financial interests the Contributor or any co-Contributors may have in companies or other entities that have an interest in the information in the Contribution or any submitted Supporting Information (e.g., grants, advisory boards, employment, consultancies, contracts, honoraria, royalties, expert testimony, partnerships, or stock ownership); and (3) indication of no such financial interests if appropriate.

Notwithstanding acceptance, the Owner or Wiley may (but is not obliged to) require changes to the Contribution, including changes to the length of the Contribution, and/or elect not to publish the Contribution if for any reason, in the Owner's or Wiley's reasonable judgment, such publication would be inconsistent with the Core Practices and associated guidelines set forth by the Committee on Publication Ethics (a not-for-profit organization based in the UK: https://publicationethics.org/core-practices) or would result in legal liability, violation of Wiley's ethical guidelines, or violation of journal ethical practices.

H. USE OF INFORMATION

The Contributor acknowledges that, during the term of this Agreement and thereafter, the Owner (and Wiley where Wiley is not the Owner) may process the Contributor's personal data, including storing or transferring data outside of the country of the Contributor's residence, in order to process transactions related to this Agreement and to communicate with the Contributor, and that the Publisher has a legitimate interest in processing the Contributor's personal data. By entering into this Agreement, the Contributor agrees to the processing of the Contributor's personal data (and, where applicable, confirms that the Contributor has obtained the permission from all other contributors to process their personal data). Wiley shall comply with all applicable laws, statutes and regulations relating to data protection and privacy and shall process such personal data in accordance with Wiley's Privacy Policy located at: https://www.wiley.com/en-us/privacy.

[X] I agree to the COPYRIGHT TRANSFER AGREEMENT as shown above, consent to execution and delivery of the Copyright Transfer Agreement electronically and agree that an electronic signature shall be given the same legal force as a handwritten signature, and have obtained written permission from all other contributors to execute this Agreement on their behalf.

Contributor's signature (type name here): Nicholas Bacci

Date: March 24, 2021

SELECT FROM OPTIONS BELOW:

[X] Contributor-owned work

[ ] U.S. Government work

Note to U.S. Government Employees

A contribution prepared by a U.S. federal government employee as part of the employee's official duties, or which is an official U.S. government publication, is called a "U.S. government work", and is in the public domain in the United States. If the Contribution was not prepared as part of the employee's duties, is not an official U.S. government publication, or if at least one author is ***not*** *a U.S. government employee, it is not a U.S. government work. If at least one author is* ***not*** *a U.S. government employee, then the non-government author should also sign the form, selecting the appropriate ownership option. If more than one author is not a U.S. government employee, one may sign on behalf of the others.*

[ ] U.K. Government work (Crown Copyright)

Note to U.K. Government Employees

The rights in a contribution prepared by an employee of a UK government department, agency or other Crown body as part of his/her official duties, or which is an official government publication, belong to the Crown and must be made available under the terms of the Open Government Licence. Contributors must ensure they comply with departmental regulations and submit the appropriate authorisation to publish. If your status as a government employee legally prevents you from signing this Agreement, please contact the Journal production editor. If this selection does not apply to at least one author in the group, this author should also sign the form, indicating transfer of those rights which that author has and selecting the appropriate additional ownership selection option. If this applies to more than one author, one may sign on behalf of the others.

[ ] Other

Including Other Government work or Non-Governmental Organisation work
Note to Non-U.S., Non-U.K. Government Employees or Non-Governmental Organisation Employees
If you are employed by the Australian Government, the World Bank, the World Health Organization, the International Monetary Fund, the European Atomic Energy Community, the Jet Propulsion Laboratory at California Institute of Technology, the Asian Development Bank, the Bank of International Settlements, USDA Agricultural Research Services, or are a Canadian Government civil servant, please download a copy of the license agreement from http://www.wileyauthors.com/licensingFAQ and upload the form to the Wiley Author Services Dashboard. If your status as a government or non-governmental organisation employee legally prevents you from signing this Agreement, please contact the Journal production editor.

Name of Government/Non-Governmental Organisation:

_____

[ ] **Company/institution owned work (made for hire in the course of employment)**
If you are an employee of Amgen, please download a copy of the company addendum from http://www.wileyauthors.com/licensingFAQ and return your signed license agreement along with the addendum. If this selection does not apply to at least one author in the group, this author should also sign the form, indicating transfer of those rights which that author has and selecting the appropriate additional ownership selection option. If this applies to more than one author, one may sign on behalf of the others.

Name of Company/Institution:

_____

Authorized Signature of Employer: _____

Date: _____

Signature of Employee: _____

Date: _____

Appendix G - Plagiarism Declaration

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

FACULTY OF HEALTH SCIENCES

PLAGIARISM DECLARATION TO BE SIGNED BY ALL HIGHER DEGREE STUDENTS

SENATE PLAGIARISM POLICY: APPENDIX ONE

I Nicholas Bacci (Student number: 472770) am a student registered for the degree of Doctor of Philosophy (PhD) in the academic year 2021.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.
- I have included as an appendix a report from "Turnitin" (or other approved plagiarism detection) software indicating the level of plagiarism in my research document.

Signature:  Date: 17/06/2021

Appendix H – Turnitin Report

a0038700:PhDThesis_NBacci_2021_PreSubmission_Turnitin.pdf

ORIGINALITY REPORT

14%	7%	11%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Nicholas Bacci, Tobias M. R. Houlton, Nanette Briers, Maryna Steyn. "Validation of forensic facial comparison by morphological analysis in photographic and CCTV samples", International Journal of Legal Medicine, 2021 Publication	6%
2	f1000research.com Internet Source	3%
3	kups.ub.uni-koeln.de Internet Source	1%
4	M. Steyn, M. Pretorius, N. Briers, N. Bacci, A. Johnson, T.M.R. Houlton. "Forensic facial comparison in South Africa: State of the science", Forensic Science International, 2018 Publication	<1%
5	www.fedoa.unina.it Internet Source	<1%
6	Félix José Fuentes Hurtado. "A system for modeling social traits in realistic faces with artificial intelligence", Universitat Politecnica de	<1%

	Valencia, 2018 Publication	
7	eprints.usq.edu.au Internet Source	<1%
8	wiredspace.wits.ac.za Internet Source	<1%
9	Submitted to University of Portsmouth Student Paper	<1%
10	fiswg.org Internet Source	<1%
11	hdl.handle.net Internet Source	<1%
12	docplayer.net Internet Source	<1%
13	Submitted to University of St. Gallen Student Paper	<1%
14	www.justice.gov Internet Source	<1%
15	www.mdpi.com Internet Source	<1%
16	Submitted to Flinders University Student Paper	<1%
17	T.M.R. Houlton, M. Steyn. "Finding Makhubu: A morphological forensic facial comparison",	<1%

	Forensic Science International, 2018 Publication	
18	Submitted to University of Hull Student Paper	<1%
19	repository.up.ac.za Internet Source	<1%
20	Submitted to Medizinischen Universität Wien Student Paper	<1%
21	amsdottorato.unibo.it Internet Source	<1%
22	Carl N. Stephan, Jodi M. Caple, Pierre Guyomarc'h, Peter Claes. "An overview of the latest developments in facial imaging", Forensic Sciences Research, 2018 Publication	<1%
23	Nicholas Bacci, Joshua Davimes, Maryna Steyn, Nanette Briers. "Development of the Wits Face Database: an African database of high-resolution facial photographs and multimodal closed-circuit television (CCTV) recordings", F1000Research, 2021 Publication	<1%
24	Submitted to University of Witwatersrand Student Paper	<1%
25	open.library.ubc.ca Internet Source	<1%