

https://dspace.org/

Forensic facial comparison: current status, limitations, and future directions.

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

2021-12-03
MDPI
https://hdl.handle.net/10539/32729
Bacci N, Davimes JG, Steyn M, Briers N. Forensic facial comparison: Current status, limitations, and future directions. Biology. 2021;10(12):1269. DOI: 10.3390/biology10121269.

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

biology

MDPI

Review

Forensic Facial Comparison: Current Status, Limitations, and Future Directions

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

Human Variation and Identification Research Unit, School of Anatomical Sciences, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg 2193, South Africa; joshua.davimes@wits.ac.za (J.G.D.); maryna.steyn@wits.ac.za (M.S.); nanette.briers@wits.ac.za (N.B.)

* Correspondence: nicholas.bacci@wits.ac.za; Tel.: +27-11-717-2204

Simple Summary: Facial identification is an emerging field in forensic anthropology, largely due to the rise in closed circuit television presence worldwide, yet there is little published research in it. Our research group has conducted a series of studies testing the validity and reliability of the facial identification practice of morphological analysis. In this paper, we summarize the results of our studies and other latest advances in facial identification practice. In addition, we present a review of relevant technical literature on the limiting factors imposed on facial identification by closed circuit television, while making recommendations for practice and the future of this research niche based on a combination of our results and the technical know-how available. Facial identification research is a multidisciplinary task, with involvement from the field of anatomy, forensic anthropology, photography, image science, and psychology, among others. The value of this brief review is the bridging of these multiple disciplines to discuss the relevant needs and requirements of facial identification in forensic practice and future research.

Abstract: Global escalation of crime has necessitated the use of digital imagery to aid the identification of perpetrators. Forensic facial comparison (FFC) is increasingly employed, often relying on poor-quality images. In the absence of standardized criteria, especially in terms of video recordings, verification of the methodology is needed. This paper addresses aspects of FFC, discussing relevant terminology, investigating the validity and reliability of the FISWG morphological feature list using a new South African database, and advising on standards for CCTV equipment. Suboptimal conditions, including poor resolution, unfavorable angle of incidence, color, and lighting, affected the accuracy of FFC. Morphological analysis of photographs, standard CCTV, and eye-level CCTV showed improved performance in a strict iteration analysis, but not when using analogue CCTV images. Therefore, both strict and lenient iterations should be conducted, but FFC must be abandoned when a strict iteration performs worse than a lenient one. This threshold ought to be applied to the specific CCTV equipment to determine its utility. Chance-corrected accuracy was the most representative measure of accuracy, as opposed to the commonly used hit rate. While the use of automated systems is increasing, trained human observer-based morphological analysis, using the FISWG feature list and an Analysis, Comparison, Evaluation, and Verification (ACE-V) approach, should be the primary method of facial comparison.

Keywords: human identification; facial identification; CCTV; photography; forensic facial comparison; morphological analysis; FISWG; face mapping; disguises

check for updates

Citation: Bacci, N.; Davimes, J.G.; Steyn, M.; Briers, N. Forensic Facial Comparison: Current Status, Limitations, and Future Directions. *Biology* **2021**, *10*, 1269. https://doi.org/10.3390/biology10121269

Academic Editors: Ann H. Ross and Eugénia Cunha

Received: 3 November 2021
Accepted: 1 December 2021
Published: 3 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Cameras and photographic imagery have been used in surveillance, identification, and detection of criminals as early as the 19th century [1]. Anthropological standards have been used to depict portraits of regular criminals for law enforcement registries, similar to today's mugshot system. These registries were intended as a means for witnesses and

victims to conduct a facial review of potential suspects. However, the lack of standardization in image capture processes made these registries ineffective. The advent of judicial photography, in the late 19th century, incorporated anthropometry and relied on standardized conditions of image capture, featuring the well-known anterior and lateral facial views with neutral expression and stance [1,2] routinely used to this day by many police departments throughout the world. The facial anthropometry application was abandoned in favor of the more accepted fingerprint identification system [3], yet the facial image capture standards it relied on endured in facial depiction practices throughout the 20th century [1].

Depicting faces [1], facial anthropometry [2], and facilitating crime scene investigations [4,5] have relied on the use of photography in a forensic context almost since its development [1]. Probably the most recognized use of photography in a forensic setting, and its derivative in the form of video recording, is surveillance. Closed-circuit television (CCTV) was the natural progression of improved use of video technology that allowed for consistent monitoring and review of potential criminal activities [6]. CCTV surveillance systems have since the 1990s become increasingly more common and relied upon throughout the world [7–10] and are in fact considered by many communities the norm in public areas [11,12].

Deployment of CCTV surveillance is considered to act as a deterrent for local crime in monitored areas [8,13,14], often shifting criminal incidents to nearby unmonitored areas instead of completely eliminating them [10]. However, perhaps its most valuable contribution is its frequent use in criminal investigations [8,15]. An analysis of CCTV data in the United Kingdom showed that when CCTV data are available, criminal activity is substantially more likely to be resolved [15]. When the data were not of use, it was primarily due to its lack of availability or some fixed parameter of the surveillance system being suboptimal, such as the incident not being covered by CCTV, the system being faulty, or the images being of insufficient quality [15]. The criteria of usefulness of CCTV recordings vary greatly based on the intended use.

Other than general surveillance and criminal activity monitoring, facial examination is often of interest for the data extracted from many CCTV surveillance systems. This has become more evident as the deployment of CCTV systems and increases in crime have led to an increase in demand for facial identification [16–18]. This rise in demand is a direct outcome of the increased availability of image data, from both CCTV data [7,16] and photographic and video evidence from other sources, such as mobile phones [19].

Forensic facial identification falls under the discipline of facial imaging, which involves the use of visual facial data to assist the identification process [20]. Through the analysis of photographic or video evidence, forensic facial identification is routinely utilized to associate persons of interest to criminal activity [17]. Craniofacial identification involves multiple disciplines, such as facial approximation, facial composites and sketches, age progression and regression, photographic superimposition, molecular photofitting, facial depiction, and facial comparison [20]. Some of these techniques, such as facial approximation and facial composites and sketches, have been researched in some depth [20]. However, forensic facial comparison (FFC) for identification remains largely untested, despite its increasing demand [17,21].

Understanding that forensic facial comparison is a niche of research that needs further development requires the use of clear terminology. A colloquial confusion in terminology between facial identification and recognition is prominent throughout many discussions. This misnomer has been discussed by Schüler and Obertová [22], who clarified that identification is reliant on perfect agreement, which is different from recognition, understood as the innate psychological process humans employ at a glance to recognize a face, usually based on familiarity. Therefore, to attempt facial identification from a forensic anthropological perspective, a strict process of facial comparison is employed. Due to the innate process of recognition in any forensic facial comparison process, the distinction needs to be made clear. Recognition is employed generally as part of the investigative process of facial com-

parison and is holistic, rapid, and methodologically inconsistent with a high predisposition to error [23,24]. Identification, however, requires further systematic analysis involving standardized, detailed, comprehensive, and meticulously recorded methodology [22]. As such, forensic facial comparison must involve the human-based detailed examination of facial images for identity confirmation [25–27].

Another prominent misconception in facial identification (ID) involves the misuse of the term "facial recognition" to specifically refer to automated or semi-automated facial recognition systems, with this being fully adopted by many in the field of automated facial recognition (e.g., [28,29]). To avoid this miscommunication, certain studies refer to automated facial recognition as facial recognition technology (FRT) or systems [30]; however, this practice is not universally applied.

The misnomer of FRT and facial ID is often closely associated to the misconception of FRT being considered the ideal approach to facial ID. FRT systems apply a variety of computer-based methods to attempt confirmation of facial identity [29,31] and have proven high levels of accuracy in constrained circumstances [28,29]. While great advances have been achieved in the field of FRT [28,32], it remains associated with high false positive rates [32,33], strong racial biases [34], and other ethical concerns around privacy and consent that require resolution prior to the employment of FRT in a legal context. Most concerns revolve around the reliance of FRT systems on biometric information [35] and highly standardized images [36–38], which are often not available in the realistic unstandardized organization of most surveillance installations. As a result, while there are strong commercial and government incentives to deploy FRT systems, in part due to their large market share (USD 3.72 billion) [39], they are still reliant on human-based validation in their operating loops [40]. The need for human validation is further enhanced by the lack of varied databases used to develop and test these FRT systems [41]. Hence, until further varied and realistic databases are used to test and develop these FRTs, human observer-based facial image comparison is considered the preferred approach to facial ID [25,42–44] and will likely persist as the validation method of choice despite the improvement and widespread deployment of FRT systems.

Understanding the limitations and permissible applications of FRTs is crucial to conducting research in both FFC and FRT. The misconceptions and assumptions around FRT and FFC may pose a risk of driving researchers and funders away from conducting research in facial identification. This is primarily because most funders and new researchers would consider facial identification, and particularly FFC, as redundant in an era where FRT has become the norm. Despite these misconceptions, human-based facial identification methods, which are currently employed routinely in the judicial system, rely on forensic facial comparison [17,42].

Facial examination, also referred to as forensic facial comparison (FFC), must be applied using the Analysis, Comparison, Evaluation, and Verification (ACE-V) approach [27], commonly used in other forensic practices, such as fingerprint identification [45]. The ACE-V methodological approach is meant to integrate principles of the scientific method in forensic comparisons in order to enhance their implementation and reliability [45].

In the past, approaches to FFC included photo-anthropometry, facial superimposition, and morphological analysis (MA) [20,27], with morphological analysis being the currently accepted method as advised by both the Facial Identification Scientific Working Group (FISWG) (https://fiswg.org/index.htm accessed on 30 October 2021) and the European Network of Forensic Science Institutes (ENFSI) (https://enfsi.eu/ accessed on 30 October 2021) [27,46]. Application of MA relies on the detailed examination of specific facial features to reach a conclusion with regard to the similarity or dissimilarity of two or more faces [27]. The facial features are assessed subjectively, evaluated, and compared between the faces [27]. The selection of individual facial features often depends on the feature list utilized. Feature lists generally include both overall face composition and structure, individual anatomical feature components (e.g., hairline shape, ear helix morphology, nasal alae protrusion, etc.), and distinguishing characteristics such as scars, blemishes, piercings,

and tattoos (e.g., [47]). The current standard feature list used for facial comparison relies on criteria developed by the FISWG for facial comparison by MA [47]. An example of how this analysis is conducted is shown in Figure 1, using sample facial images from the Wits Face Database [41].

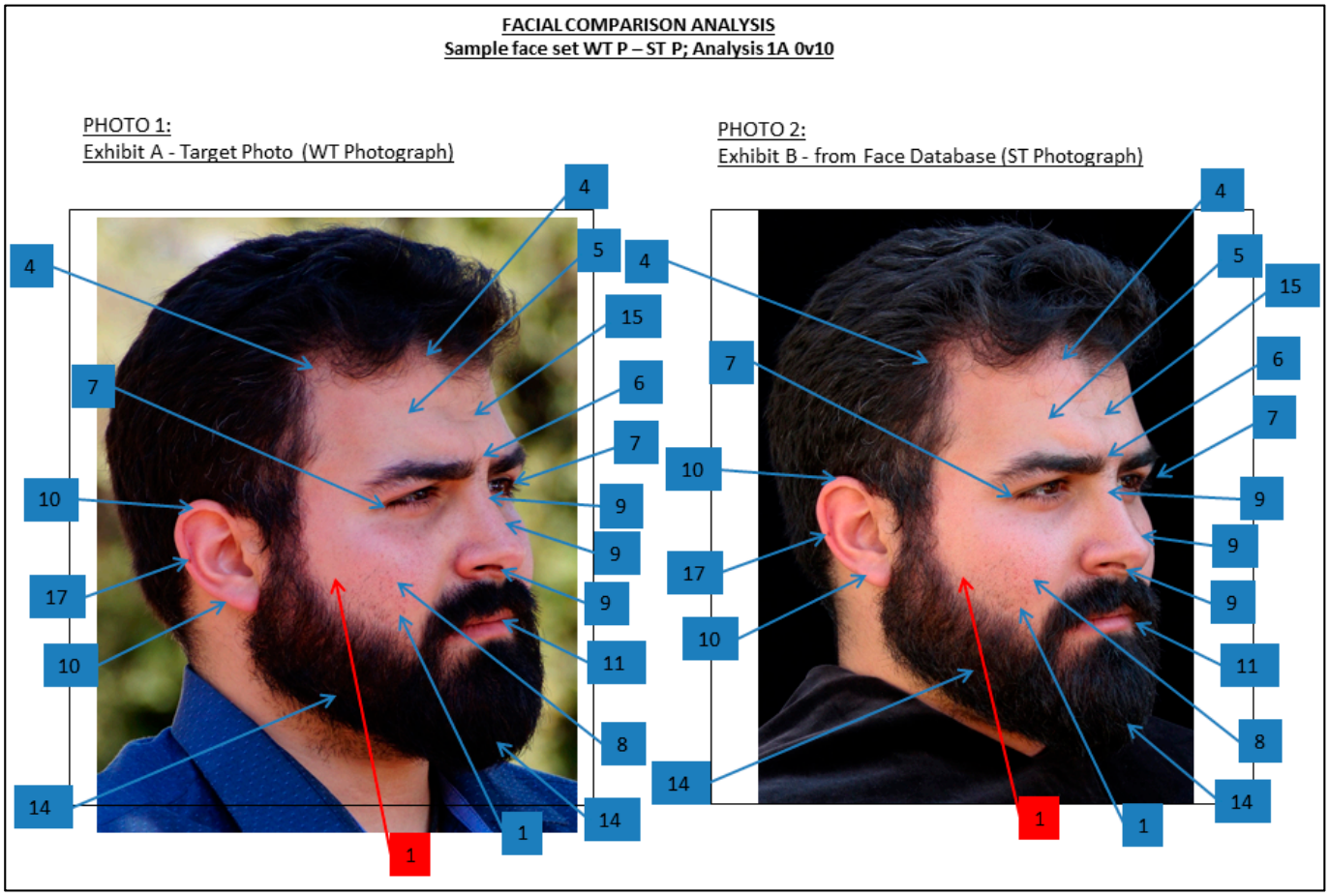


Figure 1. Example of a forensic facial comparison analysis process between a wildtype (WT) photograph and a standardized (ST) photograph from the Wits Face Database [41] sample images in the SAPS court chart format. The individual facial features are numbered, analyzed, compared, and evaluated between the two images using the FISWG feature list [47]. Features marked in blue indicate morphological similarity between the two images, while features marked in red indicate morphological dissimilarity. In the example provided, skin color appears different due to lighting discrepancies in the two images (red 1); however, skin texture appears similar (blue 1). The facial images used for Figure 1 are images of the corresponding author of the present manuscript and are part of the sample images of the Wits Face Database [41], reproducible under an open access license distributed under the terms of the Creative Commons Attribution License. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The images can be found in the Wits Face Database data note, including the supplementary material for the Wits Face Database [41].

Recently, our research group (https://www.wits.ac.za/anatomicalsciences/hviru/ accessed on 30 October 2021) has conducted a series of validation studies to test the validity and reliability of FFC using the FISWG list (https://fiswg.org/index.htm accessed on 30 October 2021) of morphological features [21,41,48,49]. The aim of this paper is to summarize the results of these findings, thus elucidating the reliability and potential uses of FFC. Potential areas of caution and observed shortcomings are also discussed. Finally, recommendations as to the minimum standards for CCTV equipment are given, as well as guidelines for future directions in research.

2. Development of an African Facial Image Database

Although various facial databases exist (e.g., [50–61]), none of these were suitable for the systematic and blind testing envisaged for the purposes of the current stream of research on FFC validation. Some of these databases have small numbers of faces (e.g., [62,63]) or contain low-resolution images (e.g., [51,64]). As these databases were developed with different purposes in mind [65], and, with the exception of one [55], do not contain African faces, a new database was needed. A database containing African faces would also be invaluable in future research on the African continent.

Such a database was developed for the purposes of these studies, but due to the magnitude of such an undertaking, currently only males are included. This new Wits Face Database includes a total of 622 unique African male individuals aged between 18 and 35 at the time of recording, each with 10 photos associated to them, in five different views (anterior, left and right lateral, and left and right 45°) [41]. The 10 photographs were captured with high-resolution midrange cameras across two different conditions: a controlled setting with uniform background and obscured clothing at a subject-to-camera distance (SCD) of 1.5 m and an uncontrolled setting with a mixed background and visible clothing at an SCD of 5 m. This brought the total to 6220 facial photographs [41]. Out of the 622 participants, 337 (54.2%) were also recorded under different CCTV recording conditions [41]. The first group, recorded under a standard digital IP CCTV installation at approximately 3 m height, included 89 individuals; the second group, recorded at an eye-level digital IP installation (1.7 m installation height), included 76 participants; the third group included 107 participants, recorded by an older analogue CCTV installation (2.5 m height); and the last group, recorded by the same digital IP CCTV camera as the first group, included 34 and 31 participants wearing caps and sunglasses, respectively [41]. Throughout the CCTV data, large amounts of data loss were experienced, particularly with the internet protocol (IP) CCTV cameras, due to corruption, compression, and intermittent connectivity (Table 1).

Table 1. Composition of the Wits Face Database [41] CCTV data and detailed data loss experienced during database development as a result of the CCTV systems' technical limitations.

Database Cohort Organization	Unique Individuals	Photographs	Corresponding CCTV [1] Recordings	Data Loss (%)
ST [2] CCTV [1]—ST [2] Photographs	98	980	89	9.2%
Eye-level CCTV [1]—ST [2] Photographs	108	1080	76	29.6%
ST [2] CCTV [1] with Cap—ST [2] Photographs	45	450	34	24.4%
ST [2] CCTV [1] with Cap—ST [2] Photographs	41	410	31	24.4%
Total IP [3] CCTV [1] Data	292	2920	230	21.2%
Analogue CCTV [1]—ST [2] Photographs	111	1110	107	3.6%
CCTV [1] Grand Totals	403	4030	337	16.4%

[1] CCTV = closed-circuit television; [2] ST = standard; [3] IP = internet protocol.

While the inclusion of males only is a good step towards expanding the diversity in populations included in face databases, the non-existence of a female database remains a notable limitation to be aware of. In principle, the FISWG feature list should be generic enough to make it applicable across sex and population groups, but facial variations may potentially lead to variations in accuracies and reliability based on the biases and abilities of the observers. The existence of a within-group face recognition advantage (previously called own- or cross-race bias) has been well described and may play a role in the reported accuracies of FFC [66–71]. It is, therefore, essential that future databases include faces that are representative of all major populations. The newly developed database is now the largest African database of CCTV recordings and matching high-resolution facial photographs. It is available for all bona fide research that meets the criteria as set out by the Human Research Ethics Committee (HREC) (Medical) of the University of the Witwatersrand [41,72].

3. Outcomes of Validation Studies

Various standards exist worldwide as to how to express the levels of confidence when it comes to possible matches. In Australia, for example, facial comparison experts are expected to present evidence strictly in descriptive terms, which can lead to suggestive language, based on the expert's prejudice and opinion [26]. In England and Wales, FFC experts report on comparisons based on the Bromby scale of support [73], where the scales of support force experts to conclude whether two compared faces are a match regardless of image conditions or quality [26]. The Bromby scale is also inherently arbitrary with no clear distinction between each step of the scale. To alleviate these uncertainties, experts from the South African Police Services (SAPS) make use of a five-point scale that reflects the ability of an expert to analyze a given set of images, as well as the confidence level of a specific conclusion [17]. For application and testing, this scale was slightly adjusted to allow statistical testing to reflect an order of severity of conclusion. Namely, a score of 1 was assigned to confident positive identifications, a score of 2 to inconclusive identifications that showed some level of morphological similarity on certain specific facial features, a score of 3 that represented an inconclusive identification with overall holistic similarity of two faces compared, a score of 4 as a negative identification, and a score of 5 indicating impossible to analyze due to insufficient visibility of landmarks [21]. A visual overview of these outcomes is shown in Figure 2.

Morphological analysis on data derived from the newly developed Wits Face Database [41,72] using the FISWG feature list [47] was found highly accurate and reliable when comparing optimal standardized photographs to wildtype (informal) unstandardized photographs [21]. In an analysis of 75 sets of faces (each containing nine no match comparisons and one positive match comparison or 10 no match comparisons—compared to a target image, total $n = 750$ comparisons), the chance corrected accuracy and reliability were found to be almost perfect in optimal photographs (99.1% and 92.1%, respectively) [21]. In the analysis of 100 face sets ($n = 1000$ comparisons) with standard digital CCTV recordings as the target image, a lower accuracy (82.6%) and reliability (74.3%) were noted [21] (Figure 2). The lower performance of MA in standard CCTV was ascribed to the variation of conditions of the different equipment and its installation. Specifically, images obtained from the standard CCTV system were of poorer quality than the high-resolution controlled and wildtype photographs, due to a number of reasons. Firstly, the image resolution of the standard digital CCTV camera was lower (4MP) than that of the photographic cameras (18MP) [21]. Secondly, the CCTV field of view was broader and less focused on the face, partly due to the SCD being approximately 3 m. As such, a larger area was captured at a lower resolution, effectively reducing the actual resolution of the recorded faces [21]. Thirdly, between the CCTV camera and the captured face, an angle of incidence of 27° was formed, which appeared to limit visibility of the face, potentially shifting relative proportions of facial features [21]. The change in perspective and the limitations it placed on the facial comparison process likely contributed to the lower accuracy and reliability seen in the standard CCTV conditions [21].

Image lighting was also markedly different between photographs and CCTV recordings, making facial characteristics reliant on color (i.e., skin tone, luminescence, and color) redundant, since they appeared different even between matching images [21]. Variations in lighting also contributed to over-exposure of certain features, effectively limiting their utility in facial comparison [21]. Beyond these discrepancies and concerns, the almost perfect accuracies and the low false positive rates identified (<1.6%) (Figure 2) are encouraging for the use of MA in a legal context from both optimal photographs and standard CCTV installations [21].

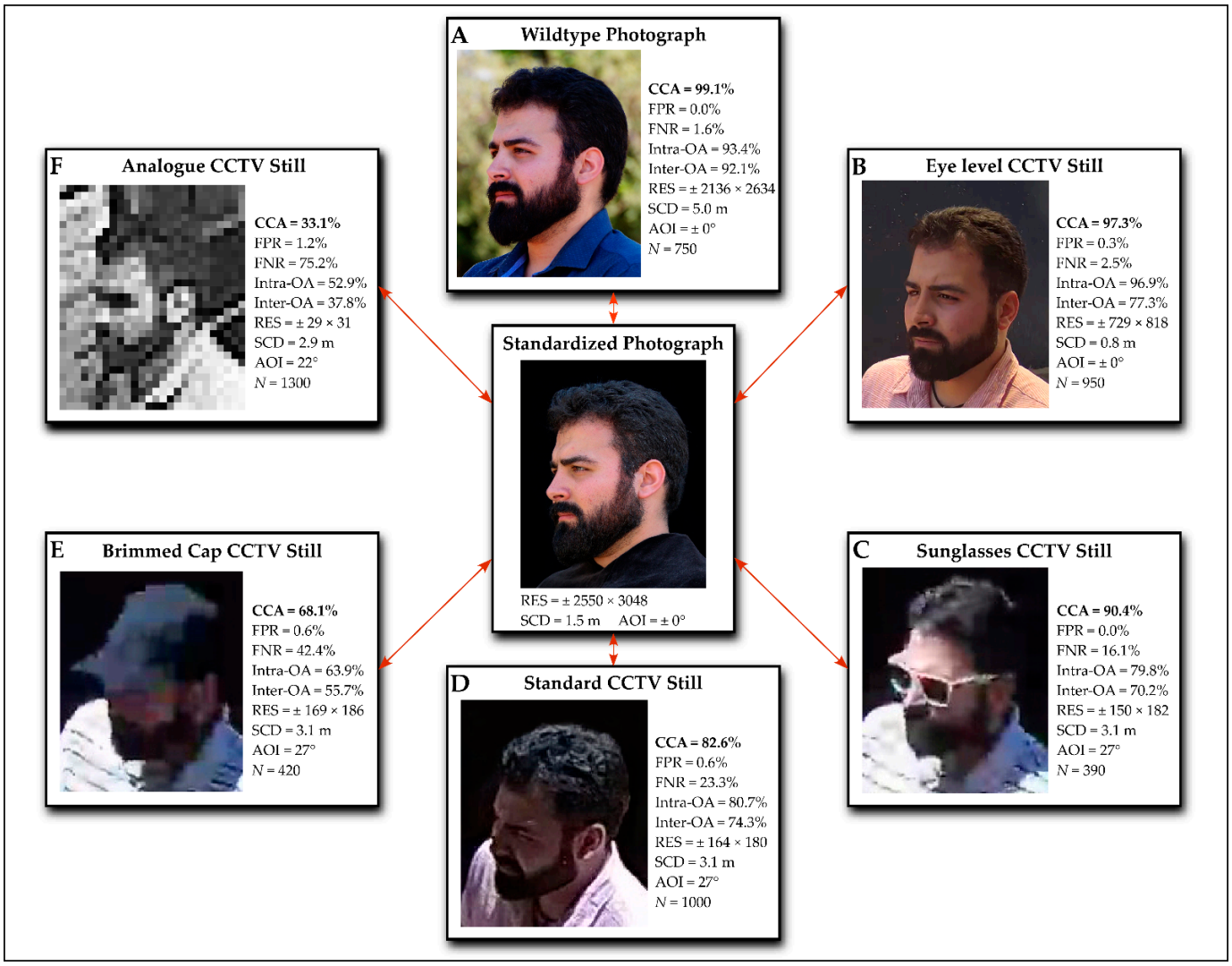


Figure 2. Visual summary of the validation studies testing morphological analysis across realistic photographic and CCTV conditions [21,48,49] using sample photographs and CCTV stills from the Wits Face Database [41]. Images (**A**) to (**F**) are samples of the target images from each set of conditions analyzed that were compared to the central image arising from the standardized photographs captured for each participant. All major statistical results and the details of the conditions of each comparison cohort are presented. Representative images of each condition are arranged from A to F in a clockwise order according to descending chance-corrected accuracy. The conditions of analysis were as follows: wildtype informal photographs (**A**) of similar quality to the standardized photographs; eye level digital CCTV still images (**B**); standard digital CCTV still images (**D**) with sunglasses (**C**) and with brimmed caps (**E**); and monochrome analogue CCTV still images (**F**). Key: CCA = chance corrected accuracy; FPR = false positive rate; FNR = false negative rate; OA = observer agreement; RES = resolution; SCD = subject-to-camera distance; AOI = angle of incidence; *N* = number of comparisons. The facial images used for Figure 2 are images of the corresponding author of the present manuscript and are part of the sample images of the Wits Face Database [41], reproducible under an open access license distributed under the terms of the Creative Commons Attribution License. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The images can be found in the Wits Face Database data note, including the supplementary material for the Wits Face Database [41].

Following on the first set of analyses under fairly optimal conditions, a second set of tests was done on 130 face sets (n = 1300 comparisons), arranged as described above, recorded on a low-resolution suboptimal analogue CCTV system. The results were found to be much poorer, with accuracies as low as 33.1% with extremely high false negative rates (75.2%) and questionable reliability (37.8%) [48] (Figure 2). The contributing factors

to this decrease in accuracy were a pronounced angle of incidence (22°), lack of color, and particularly the low-resolution images [48]. However, determining which of these specific factors contributed the most to the low accuracy is not possible by study design, but the decreased quality of the images seems to be the most problematic factor [48]. The contribution of lacking color, however, is questionable, as facial examiners in certain countries conduct their comparisons in greyscale with the consideration that color can be considered misleading. This effect of color was also observed in a previous study, where attempting to match skin color between images proved futile due to lightning discrepancies between images [21]. Irrespective of the specific contribution, the combination of these factors was highly disruptive to the facial comparison analysis—even more so than the inclusion of disguises [49].

The above suboptimal comparisons were contrasted to 95 face sets (n = 950 comparisons) recorded at eye-level with a digital IP CCTV camera. As can be expected, eye-level digital CCTV images were found to yield better results than the standard CCTV installation [21,48]. An effective 0° angle of incidence and a much smaller SCD of 0.8 m seem to have simulated the most ideal CCTV conditions for facial comparison [48]. In fact, eye-level digital CCTV recording-based facial comparisons were almost as accurate (97.3%) and reliable (77.3%) as the standardized photograph to unstandardized photograph comparisons [21,48] (Figure 2). This outcome is telling of the factors that may have played the biggest role being angle of incidence and SCD, since the standard CCTV and the eye-level CCTV were identical cameras installed at different conditions [21,48]. However, to assess the extent of the influence these factors had on facial comparison, further targeted testing of these individual factors is required.

During the analysis of the data from the facial comparisons, two iterations were conducted—the strict and the lenient iterations. Under a strict iteration, only a confident positive identification was taken as a match, while under the lenient iteration, even inconclusive analyses with some morphological similarity in facial features were considered as matches along with the positive identification [17,21]. When reviewing the performance of MA in the analogue CCTV data, it was noted that a significantly altered performance resulted under different levels of analysis strictness. A strict iteration resulted in a worse performance in the analogue CCTV comparisons than across all other comparisons (photographs, standard CCTV, and eye-level CCTV) [48]. All other analyses from the various CCTV and photographic images showed improved performance under a strict iteration [21,48,49]. This outcome advocates that under particularly suboptimal conditions, such as analogue CCTV, even a strict approach to the analysis is ineffective in improving performance. However, the decreased accuracy under a strict iteration may be worth considering as a marker of suboptimal conditions. Effectively, when a strict iteration results in lower performance of MA in a particular dataset than a lenient iteration, that dataset should be viewed as being below a usable threshold for facial comparison. As such, recordings that perform worse in a strict iteration, particularly in cases where target exclusion is not possible, should be avoided for positive identification. Effectively, when testing the performance of MA in a given dataset extracted from a specific CCTV installation, both a strict and lenient iteration should be conducted. Should the strict iteration perform worse than the lenient iteration, then the specific CCTV installation that yielded that footage should be considered below a usable threshold for the purpose of FFC. This consideration of statistical analyses is included in our recommendations on how to conduct MA.

Across all of our studies, the best measure of accuracy was found to be the chance corrected accuracy (CCA) [21,48,49]. CCA was calculated by conducting a weighted Cohen's kappa (with squared weighting) on the assigned scores for each comparison contrasted to the actual true match-up information for each comparison trial. This is different to the normal hit rate or raw accuracy, which simply indicates the amount of correctly scored trials irrespective of the degree of error or the sample composition. This is also different to the balanced accuracy that is calculated when computing a confusion

matrix analysis, which is effectively the sum of the sensitivity and specificity divided by two [74]. In the studies' results, CCA varied the most and was seen as the most representative measure of accuracy, particularly when compared to the simple hit rate and balanced accuracy. These two accuracies presented skewed results towards true negatives due to the studies being conducted under a one-to-many comparisons context, an approach to facial analysis also seen as the harshest testing criteria for automated facial recognition systems [75,76]. As a result, the non-chance corrected accuracies appeared deceptively higher due to the high prevalence of true negative matches, despite other measures of performance indicating a more questionable outcome. With this consideration, future FFC studies should consider making use of CCA as their primary measure of accuracy as opposed to hit rate, historically the most common measure of accuracy.

Beyond the optimization of surveillance system installations specifically for facial comparison, an additional limiting factor investigated in these studies were the effects of disguises on facial comparison. We specifically investigated the effect of sunglasses (n = 390) and brimmed caps (n = 420) on FFC performance [49] (Figure 2). Overall, the performance of MA in faces disguised with sunglasses was markedly high (90.4%) [49], in fact surpassing the performance of facial comparison under the same standard CCTV conditions without sunglasses (82.6%), but not better than the photographic (99.1%) [21] or eye-level CCTV data (97.3%) [48] (Figure 2). This unusual consequence of sunglasses on facial comparison has also been observed by Davis and Valentine, who tested live subject to image identification [77]. These authors [77] suggested that the instruction that was given to participants conducting face matching tasks to rely on the external facial features with subjects disguised by sunglasses apparently increased their ability to recognize a face disguised by sunglasses. External facial features are in fact considered the most reliable set of features in unfamiliar face matching, as corroborated by other studies [78,79]. In FFC, conducted using the FISWG feature list, a methodical approach with a focus on all facial features including the external ones was followed. As a result, this methodical approach may have indirectly contributed to avoiding the limitation that sunglasses would normally pose on this comparison cohort. In contrast, faces disguised by brimmed caps yielded an exceedingly low CCA (68.1%) [49], yet not nearly as low as the analogue CCTV comparisons (33.1%) [48] (Figure 2). The limitations posed by brimmed caps appeared to have been compounded by the large angle of incidence of the standard CCTV recordings as well as the strong natural lighting from the sun. These two conditions, in conjunction with the brimmed caps, created shadows over the face, obscuring an even greater number of facial features, resulting in large-scale information loss [49]. This effectively rendered comparison much more difficult, as less than the lower half of the face and the ears could be evaluated [49].

Eyeglasses and various types of hats have historically been viewed as the most inconspicuous and common disguises [25,80,81]. Although the specific effects of various disguises have been discussed broadly, only one study has attempted applying MA to a disguised sample [49]. Despite their lack of testing in MA, in face matching recognition, brimmed caps were found to increase error rates over other comparison tasks [81]. Brimless caps and glasses, on the other hand, appear to have a less pronounced effect on match accuracy, varying by the method employed [82,83].

The success of MA in disguised faces was credited in large part to the FISWG feature list [47]. The use of even rudimentary feature instructions or even partial feature lists is able to increase the performance of facial comparison analyses [84–86], with a more pronounced effect noted for trained experts [84]. Our results from the disguised test of MA [49] reinforce these outcomes of other studies, further supporting the use of feature lists in MA.

4. Discussion

This paper summarized the outcomes of our recent studies testing MA and the FISWG feature list across varied conditions of facial CCTV images and photographs. In addition, it presented and discussed the major limitations of FFC. MA of faces, using a feature list, is accurate and valid, particularly when conditions are optimal (e.g., high-resolution photographs and high-resolution CCTV with limited perspective distortion/angle of incidence). Image quality had the most notable effect on facial comparison performance (analogue CCTV recordings), while brimmed caps were found to be the second-most limiting condition. Across both of these conditions, the major limiting factor appeared to be overall loss of facial feature information, with caps obscuring almost half of the face and the poor quality of analogue video material making most of the facial details indistinguishable.

4.1. Influence of CCTV Installations

To determine the minimum criteria for facial examination across various CCTV installations, a more thorough understanding of the conditions imposed on footage by specific installations is needed. This is of particular relevance with the continuing global increase in the installation and usage of CCTV systems that has been seen across private, public, and commercial sectors in the last two decades [7,46,87,88]. This increase can be attributed to multiple factors; however, two major drivers include advancements in computing and CCTV system production and a reduction in the associated costs [7]. The vast global increase in CCTV deployment has led directly to an increase in available data for use in potential criminal surveillance and related investigations.

While this global increase in CCTV data is beneficial to criminal investigation and facial comparison, there is a concerning lack of standardization of required installation, recording conditions, and image quality [20,82,89–92]. As a result, the usefulness of CCTV-derived facial images is difficult to assess and makes facial comparison challenging in contrast to controlled photographs and mugshots. These limitations along the CCTV imaging chain are often acknowledged; however, few studies have assessed their implication in facial comparison accuracies [21,48,82,92–94]. Successful facial identification assessment is hindered by inconsistent recording conditions and poor image quality. Facial comparison accuracy and data quality are, thus, directly correlated [95,96], especially in terms of individual accuracy variation across multiple analysts [97] and individual analyst ability overestimation [98].

CCTV camera placement is one of the major limitations in terms of recording conditions. Most surveillance systems are put in place in order to monitor large crowds or entry/access points and do not have FFC in mind. The placement of the camera is based on the field of view that can be monitored and is then complemented by the mounted height above ground. Camera height relative to subject distance gives the angle of incidence, and this is an important, and often detrimental, component for extracting facial details from recordings.

Typical surveillance camera mount heights are between 2.5 and 3 m on building exteriors and ceiling height for indoor surveillance [99]. The main justification behind these mounting heights is that it lowers the risk of cameras being vandalized, stolen, or obstructed. The problem with these standardized mounting heights is that they translate to a steep angle of incidence. This in turn reduces image quality and obscures relevant facial detail as a result of the increased SCD and subsequent loss of useable resolution [48,49]. This is particularly important in facial comparison, as the amount of visible facial features and the view in which the face is seen are crucial for successful identification [100]. People also tend to naturally tilt their heads inferiorly by 15–20° when walking [100], thereby further exacerbating this problem. The current recommended angle of incidence limit is 15°, as any steeper angle would result in significant loss of facial detail [100]. Eye-level mounted cameras at 1.8 m ground height provide an approximate 0–15° angle of incidence with the subject and provide the most optimal capture of facial detail even with natural head tilt [48]. Further individual variations in facial view, or pose, are additional factors

that may further affect facial comparison, even at eye-level CCTV placement, particularly under poor quality and capture conditions [48,101,102]. Our reports [48] show an overall better and more reliable performance of MA in a digitally captured sample (IP cameras) at eye-level height (1.7 m) compared with a suboptimal sample at an angle of incidence of 27° (mount height of 3.1 m) [48]. In contrast, our analogue CCTV data, captured at an angle of incidence of 22° and height of 2.5 m, performed worse overall, although it is unclear if this was because of camera position or poor image quality [48].

Camera placement, particularly in relation to positioned angle and mounted height, dictates the monitoring area, while the camera lens, its focal length, as well as the sensor, specifically its size and number of pixels, dictate the field of view, image quality, level of optical distortion, and noise present [99]. The distance between the camera and a subject or target will then affect the image composition, which directly affects target size on sensor or picture height, and level of perspective distortion [92,103,104]. All of these factors and components will affect the usefulness of an image for facial detection and subsequent comparison analysis. Monitoring a large crowd outside a building, for example, requires 5% of picture height, while detecting a specific target requires 10% [99]. A potential target must occupy more than 400% of screen height in order to conduct facial examination, and a minimum of 1 mm must be represented per pixel of the whole image (ISO62676 recommendations) [99]. Considering the conservative European standards for facial image comparison [105], a minimum of the top quarter of a subject must be included on screen height and the face would need to represent a minimum of 1000 pixels per meter of screen height [99]. As such, for each inch (2.5 cm) of a face represented in an image, a minimum of 25.4 pixels is required [99]. For this minimum pixel density to be maintained at set SCDs, certain lens focal lengths need to be utilized. For example, at a 5 m distance from camera to subject, a focal length of 4.2 mm is necessary on a $\frac{1}{2}$" sensor HD CCTV camera [99], which is considered a common IP camera type. The longer the focal length of the lens, the narrower the field of view; simultaneously, the smaller the camera sensor, the smaller the viewing angle and the higher the noise. Bigger sensors and higher pixel counts are, in theory, always better for security and forensic applications, especially in low-light performance; however, bigger optics are then also required, which increases camera size, weight, power, and most importantly, cost.

Lighting conditions can pose further challenges in recording optimal footage. Facial details may be lost to over- or under-exposure of a subject and may not be retrievable through post-processing [99]. In outdoor locations, the position of the sun and related shadows, the amount of ambient lighting based on time of day, or the combination of multiple light sources or reflective materials near the subject or camera all could lead to unbalanced exposure. This then ties to the sensitivity of the camera sensor and its dynamic range capability. Most modern IP cameras are better suited to handling high-contrast environments, but older analogue systems generally provide either over- or under-exposed coverage with limited middle grounds [99]. Harsh and high-contrast lighting conditions often create artificial boundaries on viewed objects, altering appearances and reducing the accuracy of facial identification [99]. Over- and under-exposed footage may render an analysis impossible, based on multiple facial features being completely unrecognizable [21,106].

The capability of the camera is the primary factor in terms of low light or night-time conditions. Without the addition of directed lighting or dedicated "night-vision" cameras, CCTV systems must incorporate cameras that can record with infrared radiation (IR) and convert to visible light [107,108]. The accuracy of FFC has not been tested under IR conditions in our recent work and remains to be done in future studies. Most modern analogue and IP cameras are able to switch between day/night recording automatically and have IR LEDs built in to illuminate the target area. The range of the IR is generally limited to 20 m for midrange cameras on the market. This IR source of light could itself over-expose the subject dependent on SCD and other reflective materials present [99]. In addition, the IR footage is recorded in monochrome, and therefore,

includes the same limitations and challenges of traditional black and white CCTV footage in facial comparison [93], although in the current work, color was found to not be of much importance. Lens distortion effects and optical aberrations are more pronounced in IR cameras because of the longer wavelengths of IR [108]. Lastly, IR recording is subjected to image alterations and other artefacts based on converter quality and functioning [108]. Experimentally, IR recordings are difficult to conduct with subjects, as recording conditions need to be in low-to-zero light levels. There is a significant lack of research on IR CCTV recording in the context of facial identification and testing, and further validation of MA on a sample of IR surveillance data of comparable quality and conditions to the standard CCTV camera should be conducted.

As discussed above, placement and recording conditions of CCTV systems are crucial for reliable data capturing and use, especially in a forensic evidence context. This is inclusive of its installation in terms of network, software, and hardware. Many complications can arise as a result of these factors. Some examples experienced when attempting to develop the Wits Face Database [41,72] included inconsistent IP network connection and coverage, power outages, imminent weather problems, theft, and finally data loss, corruption, and tampering. Analogue CCTV systems for the most part do not provide remote video access and therefore require a physical storage and viewing location, limiting flexibility. These systems by default record at lower resolutions and require immediate local storage on a DVR device. This generally translates to a reduced amount of data loss and corruption compared to digital systems. Digital video can be recorded with varying rates of resolution, frame rate, and levels of compression [90]. The linkage of digital IP cameras to the internet allows for transmission of recorded footage for remote viewing, which requires high processing, storage, and data transmission capabilities. Digital video is, thus, more prone to occasional partial or complete data corruption or loss and is more perceptible to anti-forensic techniques, such as removing, hiding, and corrupting or wiping evidence from recorded footage [109–112]. In light of these threats, forensic readiness is needed in modern CCTV systems from both physical and cyber-attacks.

Little data exist describing the types and quantity of data loss incurred in CCTV systems globally and how this impacts surveillance and criminal investigations. Our studies [41] found approximately 21.2% loss of IP CCTV data and approximately 3.6% loss of analogue CCTV data during the establishment of the Wits Face Database (Table 1). CCTV data loss was noted in both IP and analogue cameras; however, the majority of corrupt or permanently lost data occurred with the digital IP camera systems. The CCTV systems utilized were an existing network at the university with no local storage and immediate transfer to a central server. During data transfer, any interruptions or fluctuations in local area network traffic or connectivity would result in data loss or irreparable corruption [41]. Studies utilizing existing CCTV systems and recordings are subjected to these types of data loss and corruption unless equipment is personally procured and installed. Data capture delays and reduced sample sizes are a considerable limitation when developing or expanding facial image databases.

The above discussed recommendations and primary limitations are generally not adhered to or considered, as is reflected in the actual data handed over to or available to law enforcement. Oftentimes, these data are of a subpar quality as a result of the numerous limitations as well as outdated camera systems [82]. Even with this subpar quality data and its limited utility, in a judiciary context, they may still successfully be implemented and should not be excluded until thoroughly reviewed first [113]. Thorough consideration of available evidence is in line with the ENFSI recommendations of triaging image data by their quality to ascertain fruitful use of FFC and efficient caseload management [114].

CCTV system installation and recording conditions are purpose driven and situationally applicable. They differ vastly to one another in terms of functionality, reliability, and environmental fit. System installation, hardware, and software need to be balanced in order to achieve the best result in terms of cost. Most systems are notably still lacking in applicability for facial comparison and are primarily disadvantaged not by installation and recording conditions but by image quality [20,82,89,90,92].

The poorer the derived image, whether it be from a photograph or CCTV footage, the lower the amount of extractable information. Image quality itself is a combination of multiple factors and related artefacts, with some of the relevant ones being resolution, pixelation, and noise. All of these are conditions that can vary notably across the various types of CCTV systems.

Analogue CCTV systems generally have lower resolutions and higher noise (grain) and often only record in monochrome. These cameras have been the global standard, and only in the last five years have we seen a large shift to internet protocol (IP) cameras [115,116]. The lower resolution leads to higher noise when attempting to enlarge the captured image for analysis and produces low clarity images [99,113].

The lack of color in most analogue recordings has a large impact in subsequent analysis, particularly in facial comparison [93]. Color plays an important role in face detection and recognition in humans, even when image quality is poor [117,118]. CCTV systems in general do not accurately capture color information from a scene [119] and have been deemed mostly unreliable in a forensic context [91,120]. Subject illumination as well as the color, orientation, and texture of objects are the primary variables dictating the accuracy of captured color information in CCTV [91]. When conducting MA using the FISWG facial feature list [47], color is the first component, and therefore, inaccurate image color data may lead to a decreased accuracy in performance. While color was easily disregarded in the majority of the analyses conducted in our studies, considering its contribution and consistency across CCTV recordings and photographs may be important for future studies.

More modern and commonly used internet protocol (IP) CCTV systems generally record full color at much higher resolutions with lower noise, as a result of high-spatial frequency blocking, overall leading to better extractable information for analysts [107]. Digital video is also a lot more flexible in recording and streaming quality compared with analogue in terms of video resolution, frame rate, and compression [90].

Poor quality CCTV recordings and extracted images have been shown to affect face matching ability in both novice and experts and leads to high overall false positive rates [82]. Image pixelation or spatial quantization, as a part of overall image quality, also drastically affects face matching ability [121]. Highly pixelated images can reduce face matching abilities by up to 50% in trained individuals when compared to a high-quality image sample [94,98]. In general, all forms of facial comparison accuracy will suffer when using low-resolution analogue CCTV images, even if image quality is good in other respects [48,81,93,122].

If we consider the SCD, the further away the subject, the greater the loss in detail in terms of representation of the face on the image. A minimum horizontal pixel count of 10–16 per face for a known face [121,123,124] and 20 pixels for an unknown face [92] is considered the bare minimum for successful identification in frontal view. Based on relative subject size on screen, Vitek et al. [125] recalculate Utochkin's [126] recommendations to 35 pixels for a known individual and 83 pixels for an unknown individual. If we are considering the effects of pixelation in a forensic setting, one needs to address the performance of matching accuracies and any form of potential enhancement, such as image blurring and reducing image size, when viewing [92]. Another important factor along the CCTV imaging chain not discussed here is that of the display fidelity and how the image is viewed on screen and the type of screen or monitor used [91].

Subject-to-camera distance can also result in facial distortion that alters facial proportions and shapes. While not exclusively investigated in the context of MA, previous work by Stephan and colleagues has looked at the SCD induced distortion and craniofacial superimposition [103,104,127]. Stephan [103] identified that discrepancy in camera-to-face, or skull, distances between photographs to be compared, presented with varying degrees of perspective distortion of facial features. At shorter distances, particularly below 6 m, the distortion was found to be more pronounced [103]. While distance-based perspective distortion is even more important in methods applying facial morphometry since at distances below 1 m a difference of 100 mm in SCD between the compared images can result in perspective distortion greater than 1% [103]. While perspective distortion would expectedly affect morphological comparison of facial features, the qualitative approach of MA and the large number of features being compared in each analysis would likely mitigate any small degrees of perspective distortion of compared images. However, further study into the effect of perspective distortion on qualitative assessment of facial features would be necessary.

The last, but important, limitation to consider in terms of image quality is video compression. As mentioned previously, digital video quality and corresponding file size can be made smaller in three ways—decreasing frame rate (e.g., 60 fps down to 5 fps); decreasing video resolution (e.g., Common image format (CIF) to Quarter CIF); and finally, by employing video compression [90,125,128]. Software video compression manipulates the spatial and temporal redundancy of moving frames in the form of CODECs, such as MPEG-4, Wavelet, H.265/HEVC, and JPEG [90,125,128]. Compression allows for large quantities of captured data to be stored in highly reduced sizes either temporarily or permanently but sacrifices image quality. Both distortion and artefacts occur when compression is introduced, hindering facial identification [90,125,129]. Keval and Sasse [90] found that the number of correct identifications of faces by untrained viewers decreased by 12–18% as MPEG-4 quality decreased and by 4–6% as Wavelet quality decreased (92–32 Kbps for both compression formats). They recommend a minimum of 52 Kbps video quality using MPEG-4 in order to achieve reliable and effective facial identification [90], albeit these results are for untrained practitioners and lower qualities would likely be reliable for trained FFC practitioners as well, perhaps not at the same magnitude. Vitek et al. [125] found correct identifications decreased from 88 to 48% as HEVC encoding quality decreased (30 kbps–15 kpbs) and they recommend 20 Kbps as a minimum threshold value. Compression employed in CCTV systems is lossy and, once performed during recording, cannot be removed or reversed. The types of distortion seen are pixelation, basis patterns, ringing, and blurring [99,130]. Recent advancements have been made improving FRT performance in light of compression artefacts; however, these artefacts remain a primary concern and drastically reduce accuracy and reliability [129]. An overview of the above-discussed various limiting factors of CCTV data in the application of MA and their specific effects in the process of facial comparison is presented in Table 2.

Table 2. Summary of CCTV systems' technical limitations in the application of morphological analysis.

General Limitations	Specific Limitations	Effects
Camera placement	• Camera height above ground [21,48,49,99,100] • Angle of incidence [21,48,49,100] • Subject-to-camera distance [103,104,127]	• Image composition affected—target size and screen/picture height [21,48,99] • Reduction of observable facial features [21,48,49,100] • Perspective distortion [103,104,127]

Table 2. *Cont.*

General Limitations	Specific Limitations	Effects
Camera specifications	• Analogue or digital [82,115,116] • Sensor size [99] • Pixel count [48,99] • Lens focal length [99]	• Reduced image quality [21,48,99] • Image distortion and artefacts [99]
Lighting conditions	• Ambient lighting [99,107,108] • Infrared vision [93,107,108]	• Loss of facial detail [48,49] • Shadows and overexposure form artificial boundaries and altered facial appearance [49,106] • Optical distortions [99]
Image quality	• Resolution [21,48] • Pixelation [48,92] • Noise/grain [99] • Video compression [90] • Color [48]	• Low clarity [99,103] • Reduced useable detail [21,48,82,90,92] • Face matching ability reduced [21,48,90,125,129]
Data loss and corruption	• Network infrastructure [41,72] • Software [99] • Hardware [41,99] • Imminent weather [41] • Power outages [41] • Compression rate [90] • Anti-forensic techniques [109–112]	• Inconsistent network connection and coverage—transfer corruption [41] • Partial or complete data loss [41] • Data tampering and removal [109–112]

In consideration of our results, only two particular CCTV camera specifications under limited conditions and installation variations were tested [21,48,49]. However, there is a large number of manufacturers that produce CCTV equipment with different specifications, requirements, and support. Testing the extent to which various market standard CCTV 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 appeared to contribute to a decrease in MA performance, particularly in an attempt to determine empirical thresholds for suitable image quality across various specifications and not only image resolution. Therefore, the common factors described above that affect quality should be investigated. For instance, developing a thorough understanding of distance-related distortion effects on MA between images from CCTV cameras and photographs could generate awareness of which features are altered more notably, and hence, increase inaccuracy at unfavorable distances. This is an important consideration for future work due to the varied conditions most CCTV systems are installed under and tailored to. The alternative of comparing faces captured under the exact same conditions would likely be more effective; however, it may not be feasible or cost-effective outside of an experimental scenario. In addition, the time discrepancy between a first set of images from a CCTV recording and a recapture for analysis 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 MA 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 the 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. However, based on the actual accuracies and the approximate sizes of faces in each CCTV setting, it would appear that when a face is composed of approximately 18 × 26 pixels or less, such as the analogue CCTV setting employed in our study [48], FFC analysis would be severely compromised. This is a suggested preliminary lower threshold as despite conditions being mostly similar between standard CCTV and analogue CCTV, in terms of angle of incidence and SCD, faces in the standard CCTV were composed of approximately double the number of pixels (41 × 52 pixels) and a much higher accuracy and reliability were obtained [21]. This threshold remains well below Vitek et al.'s [125] and Utochkin's [126] recommendations (minimum of 83 pixels for unknown faces). Developing clear, experimentally tested lowest acceptable quality thresholds, particularly under different settings and conditions, 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 [22]. Implementing this scale in conjunction with the FISWG feature list for MA could aid in identifying a threshold of confidence for the analysis process based on image quality.

Despite these uncertainties, from our earlier results, we recommend that CCTV system installations transition 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 [99]. As such, more cameras would need to be installed to cover areas previously covered by a single or pair of cameras [99]. Installing eye-level IP CCTV cameras would invariably place these systems at higher risk of vandalism and sabotage; however, the authors think this risk and increased cost are worthwhile in the context of facial comparison analyses, considering the significantly higher accuracy obtained when comparing faces recorded on these types of installations. Angle of incidence close to zero, allowing for more closely matching face views, in conjunction with high-resolution footage and the resulting quality of the facial image (at a minimum representation of a face being 41 × 52 pixels) are ideal for FFC application. While no clear benefit or shortfall of color recordings were isolated, based on the qualitative assessment of the analyses conducted, the authors would recommend the inclusion of color CCTV to allow for a wider range of feature list applications, such as the inclusion of color-based features in the FISWG feature list. However, we would also recommend the removal of color-based features as discrepancies in lighting are common between realistic recordings and ideal photographs captured for comparison. The resulting analysis of facial feature descriptors relying on color, or other factors that can vary easily and unknowingly, such as luminescence, in response to slight variations in lighting conditions should be reconsidered or removed from feature lists, as they were found either unreliable or unusable in most comparisons.

4.2. Feature List Usage, Disguises, and Training

Both the FISWG and the ENFSI recommend MA as the best practice for forensic facial identification [27,125]. In addition, FISWG advises against the use of photo-anthropometry for facial image comparison and recommends superimposition only to be utilized in conjunction with MA [27]. FISWG developed and made freely available an extensive facial feature list for use in MA [51,54]. This list includes 18 facial components, each with associated descriptors, as well as a nineteenth descriptive component for use with uncategorized features [51,54]. The FISWG feature list is also the most exhaustive list available, including over 130 facial component characteristics and over 290 characteristic descriptors [51].

The application of facial feature lists, such as the FISWG one, is different from previous feature-based comparison methods that involved facial feature classification schemes. These classification schemes were used as a way for an analyst to score each facial feature into categories based on descriptive qualities (e.g., pointed chin, broad nose bridge, etc.) [59]. The FISWG approach instead expects a facial analyst to subjectively describe the compared faces by providing an extensive list of features and descriptors to use in order to make statements based on similarities and dissimilarities [27,51]. This descriptive approach is preferred as classification schemes are viewed as prone to high inter-observer error [55,59,109]. In addition, for classification schemes to be effective, they need to be tailored to specific populations, which has only been considered by two studies to date [58,127]. Population homogeneity, however, can be problematic for classification schemes, since high prevalence of a feature classification in a given population could result in an overlapping score, leading to erroneous false positive matches [60]. On the other hand, classification schemes may be too restrictive and make scoring near impossible under certain circumstances [26]. In this series of studies, the FISWG feature list was found to greatly aid both in the training of the analyst as well as during the analysis process to achieve mostly high accuracies and good reliability levels with the exception of the lowest quality of CCTV recordings. The feature list was also found to be applicable to African male faces due to its descriptive nature, as opposed to population-specific classification schemes. Certain descriptors were found cumbersome to utilize; for example, as mentioned above, skin color and luminance were often ignored due to a mismatch, despite confirmation that two faces were indeed the same. A revision of some of these descriptors would be required to optimize the analysis process and applicability of the FISWG feature list to a broader number of settings and CCTV conditions.

In addition, while the importance of a feature list in MA is undeniable, combining a systematic approach and a feature list must involve the option to discard potential dissimilarities when it is justifiable to do so. This is possible with large feature lists such as the FISWG one, which allows for exclusion of questionable or hard to analyze features. A smaller feature list be employed would compromise the exclusion of dissimilar features that could be justified as dissimilar due to image conditions, leading to the false exclusion of a positive face match due to features varying under the different image conditions. To this end, a threshold of the number of minimum features required to conduct an analysis should be investigated as neither the feature list [47] nor the concluding statements [17] provided one.

A further consideration to improve the applicability of the FISWG and any other feature lists would be to develop specific criteria to be applied for comparisons under different disguised or obstructed faces. Once established, these criteria could be included in analyst training to prioritize features by type of disguise. This approach would be applicable in settings where facial features may not be 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 the use of face masks currently due to the spread of COVID-19. Face masks, which can vary in shape, size, and the resulting proportion of the face covered, have been shown to reduce automated facial recognition performance by 5 to 50% depending on the specific algorithm and extent of the face covered [76]. The deleterious effect in performance seemed to vary based on the color and shape of the masks as well [76]. It would, hence, be crucial to consider face masks in further tests of MA under disguised conditions as their impact on human observer-based facial comparison has not yet been considered.

While the current studies validated MA when the FISWG guidelines [27,47] were used, an array of further studies is required. 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 FFC by MA. Our recommended method of conducting MA is shown in Figure 3. This approach outlines the stepwise process of applying the recommended image quality triage by the ENFSI [114] to FISWG's ACE-V application [27] of the FISWG feature list [47] and then subdivided by intended use in a research or judicial context. Based on the application context, different approaches are used for verification; in addition, research use of MA would require further statistical analyses (Figure 3).

Based on the outcomes of our group's studies, expanding current training programs and developing new ones to increase the competence of facial comparison experts will be crucial for consistent and reliable application of FFC. The application of a feature list and an ACE-V approach by members of the public in forensic facial comparisons is not sufficient to achieve expertise. Training experts with the explicit role to conduct FFC analyses, with the use of the FISWG feature list and an ACE-V approach, is of utmost importance in a judicial context [131]. The role of expertise is particularly relevant in FFC, since unfamiliar face matching is considered complex and unreliable on all accounts [97,122]. Experts, in fact, perform notably better than members of the public [42], even when image quality was taken into consideration [132]. This expertise undoubtably arises from training in the nuances of faces, such as facial expressions and ageing changes, and acceptable anatomical variations and image-based variations between faces, beyond just the inclusion of the use of a feature list. The need for adequate training of all FFC practitioners is vital to the good standing of the practice and its admissibility in a legal context. Particularly when considering that certain countries may experience a shortage of expertise and heavy caseloads, such as South Africa, where only 30 trained specialists in the entirety of the national police force are trained to conduct FFC and testify in court to defend their conclusions [17].

The FISWG has put forward a document describing guidelines for training and expertise requirements of FFC analysts and trainers [133]. While these guidelines are crucial to the development of training courses, to the authors' knowledge, no formal standardized training or certification platforms exist for FFC [17]. A recent study on the performance of informal training courses on facial comparison suggested that there are large discrepancies between courses in improvement of facial examiner expertise [134]. We hope our recommended stepwise process to the applications of MA (Figure 3) will aid in streamlining both MA training and application. Recently, members of our research group proposed an outline for a training course with a three-tiered approach offered to the police force [17]; it inadvertently follows most of the proposed guidelines from FISWG. The first tier of training involves developing basic background knowledge of facial anatomy, evidence evaluation, image science, facial recognition psychology, and court proceedings, among other topics [17]. The second tier involves training in detailed MA using the FISWG standards and developing court-ready reports and charts [17]. The third tier is a national specific tier that involves advanced training in court proceedings and evidence presentation as well as troubleshooting from past casework in order to also train experienced peer reviewers [17], who are vital to the ACE-V application of MA. While this approach to training has not been experimentally tested, the trained police members have found success in their roles as facial examiners.

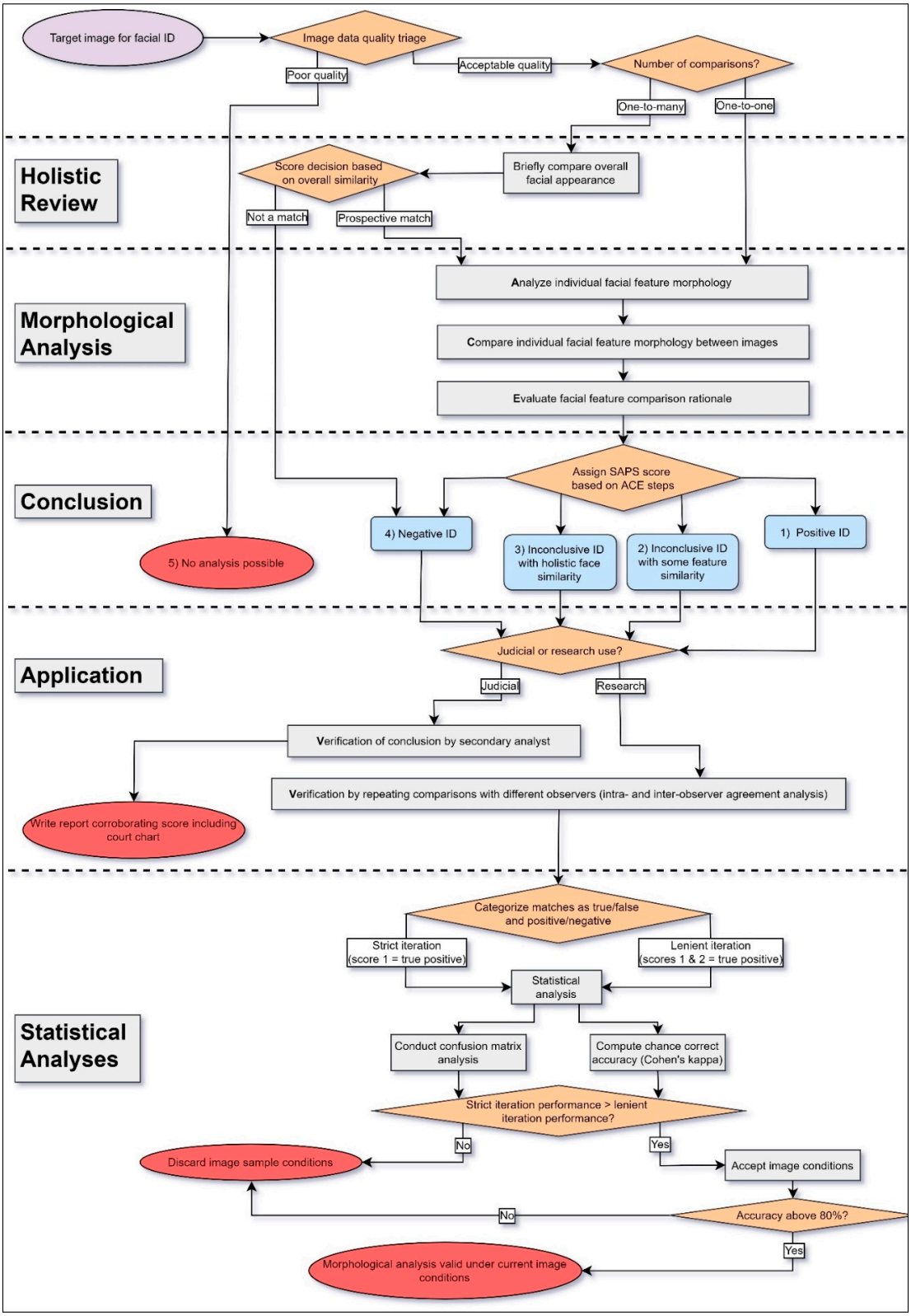


Figure 3. Flow diagram of the recommended morphological analysis process. This approach to morphological analysis uses an ACE-V method in conjunction with the FISWG feature list [47], with the inclusion of the ENFSI's image quality triaging [114] and the use of the South African Police Services (SAPS) scoring criteria [17] as adapted for research application [21]. Statistical analyses for research use are also recommended based on our recent work [48] to allow for more detailed result interpretation and comparison among future studies.

5. Conclusions

The outcomes and recommendations arising from these studies should be considered under the limitations of the investigative approach deployed. These studies attempted to simulate real-world conditions, with its array of limitations, in a select number of scenarios. These scenarios included ideal comparable photographic images, standard digital CCTV, eye-level installation digital CCTV, standard 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 standardization required for experimentation. In doing so, although realistic, the conditions were limited to the major questions broadly investigated by each study [21,48,49]. 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, SCD, distortions, color, visibility of features, and more. In testing the specific conditions outlined in each of the above studies, controlling for or identifying which of the multiple limiting factors contributed to the poor performance of MA would be impossible. Even when within likely tolerable degrees, these limiting factors cannot be isolated from one another in certain circumstances. However, with this baseline of conditions and considerations, future studies can be tailored to the specific limitations that CCTV imposes on FFC in a highly controlled setting to determine the exact contribution of each of these limiting factors to the accuracy of MA.

With these concerns and limitations clearly stated, future studies should be focused to target specific limiting factors individually in order to develop a clear threshold for image data to be usable for facial comparison. While other approaches to facial identification as a whole may also be gaining popularity, such as the increasing performance of automated systems [31,75,76] and the deployment of super-recognizers [135,136], continued research in forensic facial comparison by MA is crucial as the most universally applicable and reliable method. The importance of MA-based FFC is especially noteworthy in law enforcement applications, where the majority of available image data is of low to poor quality [113]. As such, the authors strongly advise that trained human observer-based MA, using the FISWG feature list [47] and an ACE-V approach [27], should remain the principal method of facial comparison for identification purposes, as recommended by both the FISWG and ENFSI [27,114].

Author Contributions: Conceptualization, N.B. (Nicholas Bacci), M.S. and N.B. (Nanette Briers); methodology, N.B. (Nicholas Bacci) and J.G.D.; formal analysis, N.B. (Nicholas Bacci); investigation, N.B. (Nicholas Bacci); visualization, N.B. (Nicholas Bacci) and J.G.D.; resources, M.S. and N.B. (Nanette Briers); data curation, N.B. (Nicholas Bacci) and J.G.D.; writing—original draft preparation, N.B. (Nicholas Bacci), J.G.D. and N.B. (Nanette Briers); writing—review and editing, N.B. (Nicholas Bacci), J.G.D., M.S. and N.B. (Nanette Briers); supervision, M.S. and N.B. (Nanette Briers); project administration, N.B. (Nicholas Bacci); funding acquisition, N.B. (Nicholas Bacci), M.S. and N.B. (Nanette Briers). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the South African National Research Foundation (NRF), DAAD-NRF Bursary, and the J.J.J. Smieszek Fellowship from the School of Anatomical Sciences, University of the Witwatersrand awarded to N. Bacci (Grant No.: 11858) and NRF funds awarded to N. Briers as part of the Improving Methodologies and Practices in Cranio-facial 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 were not involved in the study design, data collection, analysis, interpretation, or writing of the manuscript or submission. The APC was funded by N. Briers' and M. Steyn's Research Incentive (RINC) funds.

Institutional Review Board Statement: The core studies [21,48,49] discussed in the present review paper were conducted according to the guidelines of the Declaration of Helsinki and approved by the Human Research Ethics Committee (Medical) of the University of the Witwatersrand, Johannesburg. The ethics approval number is M171026, and approval was granted on 13 March 2018.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the core studies [21,48,49] discussed in the present review paper. Written informed consent has been obtained from the corresponding author to publish his facial images as part of this paper.

Data Availability Statement: The facial image data used in Figures 1 and 2 for the current paper are of the corresponding author and publicly available as part of the sample images of the Wits Face Database [41], reproducible under an open access license distributed under the terms of the Creative Commons Attribution License. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. These images can be accessed as part of the Wits Face Database data note, via the supplementary material for the Wits Face Database [41].

Acknowledgments: Thanks are due to the support provided from the South African National Research Foundation (NRF), DAAD-NRF Bursary, and the J.J.J. Smieszek Fellowship from the School of Anatomical Sciences, University of the Witwatersrand. DAAD-NRF and J.J.J. Smieszek Fellowship funds were awarded to N. Bacci (Grant No.: 11858) and NRF funds were awarded to N. Briers as part of the Improving Methodologies and Practices in Cranio-facial 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 were not involved in study design, data collection, analysis, interpretation, or writing of the manuscript or its submission. Acknowledgement is also owed to all the participants who agreed to be photographed and recorded for the development of the Wits Face Database. Thanks are due to 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. Gratitude is also owed to Tobias Houlton for his aid in study validation. Lastly, thanks are due to the volunteers who aided in participant recruitment for the Wits Face Database: 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, and Jessica Manavhela.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Jäger, J. Photography: A means of surveillance? Judicial photography, 1850 to 1900. *Crime Hist. Sociétés* **2001**, *5*, 27–51. [CrossRef]
2. Bertillon, A.; McClaughry, R.W. *Signaletic Instructions Including the Theory and Practice of Anthropometrical Identification*; McClaughry, R.W., Ed.; The Werner Company: Chicago, IL, USA, 1896.
3. Faigman, D.L. Anecdotal Forensics, Phrenology, and Other Abject Lessons from the History of Science. *Hastings Law J.* **2008**, *59*, 979–1000.
4. Mokwena, R.J. *The Value of Photography in the Investigation of Crime Scenes*; University of South Africa: Pretoria, South Africa, 2012.
5. Bell, A. Crime scene photography in England, 1895–1960. *J. Br. Stud.* **2018**, *57*, 53–78. [CrossRef]
6. Lindegaard, M.R.; Bernasco, W. Lessons Learned from Crime Caught on Camera. *J. Res. Crime Delinq.* **2018**, *55*, 155–186. [CrossRef] [PubMed]
7. 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*, 110–135. [CrossRef]
8. Piza, E.L.; Welsh, B.C.; Farrington, D.P.; Thomas, A.L. CCTV surveillance for crime prevention: A 40-year systematic review with meta-analysis. *Criminol. Public Policy* **2019**, *18*, 135–159. [CrossRef]
9. Jain, A.K.; Klare, B.; Park, U. Face Matching and Retrieval in Forensics Applications. *IEEE Multimed.* **2012**, *19*, 20. [CrossRef]
10. Moyo, S. *Evaluating the Use of CCTV Surveillance Systems for Crime Cotnrol and Prevention: Selected Case Studies from Johannesburg and Tshwane, Gauteng*; University of South Africa: Pretoria, South Africa, 2019.
11. Goold, B.; Loader, I.; Thumala, A. The banality of security: The curious case of surveillance cameras. *Br. J. Criminol.* **2013**, *53*, 977–996. [CrossRef]
12. Duncan, J. How CCTV surveillance poses a threat to privacy in South Africa. *Conversation* **2018**, 1–3. Available online: https://theconversation.com/how-cctv-surveillance-poses-a-threat-to-privacy-in-south-africa-97418 (accessed on 30 October 2021).
13. Gill, M.; Spriggs, A. Assessing the Impact of CCTV: Home Office Research Study 292. 2005. Available online: https://techfak.uni-bielefeld.de/~{}iluetkeb/2006/surveillance/paper/social_effect/CCTV_report.pdf (accessed on 25 November 2021).

14. Welsh, B.C.; Farrington, D.P. Effects of Closed Circuit Television Surveillance on Crime. *Campbell Syst. Rev.* **2008**, *4*, 1–73. [CrossRef]
15. Ashby, M.P.J. The Value of CCTV Surveillance Cameras as an Investigative Tool: An Empirical Analysis. *Eur. J. Crim. Policy Res.* **2017**, *23*, 441–459. [CrossRef]
16. Kleinberg, K.F.; Siebert, J.P. A study of quantitative comparisons of photographs and video images based on landmark derived feature vectors. *Forensic Sci. Int.* **2012**, *219*, 248–258. [CrossRef]
17. Steyn, M.; Pretorius, M.; Briers, N.; Bacci, N.; Johnson, A.; Houlton, T.M.R. Forensic facial comparison in South Africa: State of the science. *Forensic Sci. Int.* **2018**, *287*, 190–194. [CrossRef]
18. Jackson, A. The Admissibility of Identification Evidence Made on the Basis of Recognition from Photographs Taken at a Crime Scene. *J. Crim. Law* **2016**, *80*, 234–236. [CrossRef]
19. Houlton, T.M.R.; Steyn, M. Finding Makhubu: A morphological forensic facial comparison. *Forensic Sci. Int.* **2018**, *285*, 13–20. [CrossRef]
20. Stephan, C.N.; Caple, J.M.; Guyomarc'h, P.; Claes, P. An overview of the latest developments in facial imaging. *Forensic Sci. Res.* **2019**, *4*, 10–28. [CrossRef] [PubMed]
21. Bacci, N.; Houlton, T.M.R.; Briers, N.; Steyn, M. Validation of forensic facial comparison by morphological analysis in photographic and CCTV samples. *Int. J. Legal Med.* **2021**, *135*, 1965–1981. [CrossRef] [PubMed]
22. Schüler, G.; Obertová, Z. Visual identification of persons: Facial image comparison and morphological comparative analysis. In *Statistics and Probability in Forensic Anthropology*; Obertová, Z., Stewart, A., Cattaneo, C., Eds.; Elsevier Academic Press: London, UK, 2020; pp. 313–330.
23. Behrman, B.W.; Davey, S.L. Eyewitness identification in actual criminal cases: An archival analysis. *Law Hum. Behav.* **2001**, *25*, 475–491. [CrossRef]
24. Boyce, M.A.; Lindsay, D.S.; Brimacombe, C.A.E. Investigating investigators: Examining the impact of eyewitness identification evidence on student-investigators. *Law Hum. Behav.* **2008**, *32*, 439–453. [CrossRef] [PubMed]
25. Davis, J.P.; Valentine, T.; Wilkinson, C. Facial image comparison. In *Craniofacial Identification*; Wilkinson, C., Rynn, C., Eds.; Cambridge University Press: New York, NY, USA, 2012; pp. 136–153. ISBN 9781139049566.
26. Valentine, T.; Davis, J.P. *Forensic Facial Identification: Theory and Practice of Identification from Eyewitnesses, Composites and CCTV*; John Wiley & Sons, Ltd.: Chichester, UK, 2015; ISBN 9781118469538.
27. Facial Identification Scientific Working Group. Facial Comparison Overview and Methodology Guidelines. 2019. Available online: https://fiswg.org/fiswg_facial_comparison_overview_and_methodology_guidelines_V1.0_20191025.pdf (accessed on 25 November 2021).
28. Adjabi, I.; Ouahabi, A.; Benzaoui, A.; Taleb-Ahmed, A. Past, present, and future of face recognition: A review. *Electronics* **2020**, *9*, 1188. [CrossRef]
29. Akhtar, Z.; Rattani, A. A Face in any Form: New Challenges and Opportunities for Face Recognition Technology. *Computer* **2017**, *50*, 80–90. [CrossRef]
30. Lai, X.; Patrick Rau, P.L. Has facial recognition technology been misused? A user perception model of facial recognition scenarios. *Comput. Hum. Behav.* **2021**, *124*, 106894. [CrossRef]
31. Kaur, P.; Krishan, K.; Sharma, S.K.; Kanchan, T. Facial-recognition algorithms: A literature review. *Med. Sci. Law* **2020**, *60*, 131–139. [CrossRef] [PubMed]
32. Grother, P.; Ngan, M.; Hanaoka, K. *Face Recognition Vendor Test (FRVT) Part 2: Identification*; US Department of Commerce, National Institute of Standards & Technology: Gaithersburg, MD, USA, 2019.
33. Dodd, V. UK Police Use of Facial Recognition Technology a Failure, Says Report. *The Guardian*, 15 May 2018.
34. Grother, P.; Ngan, M.; Hanaoka, K. *Face Recognition Vendor Test Part 3: Demographic Effects*; US Department of Commerce, National Institute of Standards & Technology: Gaithersburg, MD, USA, December 2019.
35. White, D.; Jonathon Phillips, P.; Hahn, C.A.; Hill, M.; O'Toole, A.J. Perceptual expertise in forensic facial image comparison. *Proc. R. Soc. B Biol. Sci.* **2015**, *282*, 20151292. [CrossRef]
36. Davis, J.P.; Valentine, T.; Davis, R.E. Computer assisted photo-anthropometric analyses of full-face and profile facial images. *Forensic Sci. Int.* **2010**, *200*, 165–176. [CrossRef] [PubMed]
37. Urbanová, P. Performance of distance-based matching algorithms in 3D facial identification. *Egypt. J. Forensic Sci.* **2016**, *6*, 135–151. [CrossRef]
38. Press Association. Welsh Police Wrongly Identify Thousands as Potential Criminals. *The Guardian*, 5 May 2018.
39. Mordor Intelligence. Global Facial Recognition Market (2021–2026). 2020. Available online: https://www.mordorintelligence.com/industry-reports/facial-recognition-market (accessed on 30 October 2021).
40. Spaun, N.A. Facial comparisons by subject matter experts: Their role in biometrics and their training. *Int. Conf. Biom.* **2009**, *5558*, 161–168. [CrossRef]
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**, *10*, 131. [CrossRef]
42. 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*, 191–196. [CrossRef]

43. Valentine, T.; Davis, J.P. *Forensic Facial Identification*; John Wiley & Sons: Hoboken, NJ, USA, 2015; ISBN 9781118469118.
44. White, D.; Dunn, J.D.; Schmid, A.C.; Kemp, R.I. Error Rates in Users of Automatic Face Recognition Software. *PLoS ONE* **2015**, *10*, e0139827. [CrossRef]
45. Speckeis, C. Can ACE-V be validated? *J. Forensic Identif.* **2011**, *61*, 201–209.
46. Norris, C. The success of failure. Accounting for the global growth of CCTV. In *Routledge Handbook of Surveillance Studies*; Ball, K., Haggerty, K.D., Lyon, D., Eds.; Routledge: London, UK; New York, NY, USA, 2012; pp. 251–258.
47. Facial Identification Scientific Working Group. Facial Image Comparison Feature List for Morphological Analysis. 2018. Available online: https://fiswg.org/FISWG_Morph_Analysis_Feature_List_v2.0_20180911.pdf (accessed on 25 November 2021).
48. Bacci, N.; Steyn, M.; Briers, N. Performance of forensic facial comparison by morphological analysis across optimal and suboptimal CCTV settings. *Sci. Justice* **2021**, *61*, 743–754. [CrossRef]
49. Bacci, N.; Briers, N.; Steyn, M. Assessing the effect of facial disguises on forensic facial comparison by morphological analysis. *J. Forensic Sci.* **2021**, *66*, 1220–1233. [CrossRef]
50. Martinez, A.M.; Benavente, R. *The AR Face Database CVC Technical Report #24*; Universitat Autònoma de Barcelona: Bellaterra, Spain, 1998; Volume 24.
51. Phillips, P.J.; Moon, H.; Rizvi, S.A.; Rauss, P.J. The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1090–1104. [CrossRef]
52. Sim, T.; Baker, S.; Bsat, M. The CMU Pose, Illumination, and Expression (PIE) database. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 1615–1618.
53. Calvo, M.G.; Lundqvist, D. Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behav. Res. Methods* **2008**, *40*, 109–115. [CrossRef] [PubMed]
54. Gao, W.; Cao, B.; Shan, S.; Chen, X.; Zhou, D.; Zhang, X.; Zhao, D. The CAS-PEAL large-scale chinese face database and baseline evaluations. *IEEE Trans. Syst. Man Cybern.* **2008**, *38*, 149–161.
55. Milborrow, S.; Morkel, J.; Nicolls, F. The MUCT Landmarked Face Database. *Pattern Recognit. Assoc. S. Afr.* **2008**. Available online: http://www.milbo.org/muct/ (accessed on 25 November 2021).
56. Tottenham, N.; Tanaka, J.W.; Leon, A.C.; McCarry, T.; Nurse, M.; Hare, T.A.; Marcus, D.J.; Westerlund, A.; Casey, B.J.; Nelson, C. The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Res.* **2009**, *168*, 242–249. [CrossRef]
57. Grgic, M.; Delac, K.; Grgic, S. SCface—Surveillance cameras face database. *Multimed. Tools Appl.* **2011**, *51*, 863–879. [CrossRef]
58. Kostinger, M.; Wohlhart, P.; Roth, P.M.; Bischof, H. Annotated Facial Landmarks in the Wild: A large-scale, real-world database for facial landmark localization. In Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), Barcelona, Spain, 6–13 November 2011; IEEE: Barcelona, Spain, 2011; pp. 2144–2151.
59. Sagonas, C.; Antonakos, E.; Tzimiropoulos, G.; Zafeiriou, S.; Pantic, M. 300 Faces In-The-Wild Challenge: Database and results. *Image Vis. Comput.* **2015**, *47*, 3–18. [CrossRef]
60. Karam, L.J.; 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*, 93940. [CrossRef]
61. Ma, D.S.; Correll, J.; Wittenbrink, B. The Chicago face database: A free stimulus set of faces and norming data. *Behav. Res. Methods* **2015**, *47*, 1122–1135. [CrossRef]
62. Belhumeur, P.N.; Kriegman, D.J. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 711–720. [CrossRef]
63. Georghiades, A.S.; Belhumeur, P.N.; Kriegman, D.J. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 643–660. [CrossRef]
64. Huang, G.B.; Ramesh, M.; Berg, T.; Learned-Miller, E. *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*; University of Massachusetts: Amherst, MA, USA, 2007; Volume October, pp. 1–11.
65. Gross, R. Face Databases. In *Handbook Face Recognition*; Springer: New York, NY, USA, 2005; pp. 301–327. [CrossRef]
66. Shapiro, P.N.; Penrod, S. Meta-analysis of facial identification studies. *Psychol. Bull.* **1986**, *100*, 139–156. [CrossRef]
67. Bothwell, R.K.; Brigham, J.C.; Malpass, R.S. Cross-Racial Identification. *Pers. Soc. Psychol. Bull.* **1989**, *15*, 19–25. [CrossRef]
68. Anthony, T.; Copper, C.; Mullen, B. Cross-Racial Facial Identification: A Social Cognitive Integration. *Pers. Soc. Psychol. Bull.* **1992**, *18*, 296–301. [CrossRef]
69. Meissner, C.A.; Brigham, J.C. Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychol. Public Policy Law* **2001**, *7*, 3–35. [CrossRef]
70. Chiroro, P.M.; Tredoux, C.G.; Radaelli, S.; Meissner, C.A. Recognizing faces across continents: The effect of within-race variations on the own-race bias in face recognition. *Psychon. Bull. Rev.* **2008**, *15*, 1089–1092. [CrossRef] [PubMed]
71. Kokje, E.; Bindemann, M.; Megreya, A.M. Cross-race correlations in the abilities to match unfamiliar faces. *Acta Psychol.* **2018**, *185*, 13–21. [CrossRef]
72. Bacci, N.; Davimes, J.; Steyn, M.; Briers, N. Wits Face Database. Available online: https://hdl.handle.net/10539/29924 (accessed on 30 October 2021).
73. Bromby, M. CCTV and Expert Evidence: Addressing the Reliability of New Sciences. *Archbold News*, 2 November 2006; 6–9.
74. Kuhn, M. Caret: Classification and Regression Training R Package. 2020. Available online: https://cran.r-project.org/web/packages/caret/index.html (accessed on 30 October 2021).

75. Ngan, M.; Grother, P.; Hanaoka, K. *Ongoing Face Recognition Vendor Test (FRVT) Part 6A: Face Recognition Accuracy with Masks Using Pre-COVID-19 Algorithms*; US Department of Commerce, National Institute of Standards & Technology: Gaithersburg, MD, USA, 2020.
76. Ngan, M.; Grother, P.; Hanaoka, K. *Ongoing Face Recognition Vendor Test (FRVT) Part 6B: Face Recognition Accuracy with Face Masks Using Post-COVID-19 Algorithms*; US Department of Commerce, National Institute of Standards & Technology: Gaithersburg, MD, USA, 2020.
77. Davis, J.P.; Valentine, T. CCTV on trial: Matching video images with the defendant in the dock. *Appl. Cogn. Psychol.* **2009**, *23*, 482–505. [CrossRef]
78. Bruce, V.; Henderson, Z.; Greenwood, K.; Hancock, P.J.B.; Burton, A.M.; Miller, P. Verification of face identities from images captured on video. *J. Exp. Psychol. Appl.* **1999**, *5*, 339–360. [CrossRef]
79. Kemp, R.I.; Caon, A.; Howard, M.; Brooks, K.R. Improving Unfamiliar Face Matching by Masking the External Facial Features. *Appl. Cogn. Psychol.* **2016**, *30*, 622–627. [CrossRef]
80. Terry, R.L. How wearing eyeglasses affects facial recognition. *Curr. Psychol.* **1993**, *12*, 151–162. [CrossRef]
81. Henderson, Z.; Bruce, V.; Burton, A.M. Matching the faces of robbers captured on video. *Appl. Cogn. Psychol.* **2001**, *15*, 445–464. [CrossRef]
82. 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. CAHId* **2009**, *1*, 19–28.
83. Righi, G.; Peissig, J.J.; Tarr, M.J. Recognizing disguised faces. *Vis. Cogn.* **2012**, *20*, 143–169. [CrossRef]
84. Towler, A.; White, D.; Kemp, R.I. Evaluating the feature comparison strategy for forensic face identification. *J. Exp. Psychol. Appl.* **2017**, *23*, 47–58. [CrossRef] [PubMed]
85. Megreya, A.M.; Bindemann, M. Feature instructions improve face-matching accuracy. *PLoS ONE* **2018**, *13*, e0193455. [CrossRef] [PubMed]
86. Megreya, A.M. Feature-by-feature comparison and holistic processing in unfamiliar face matching. *PeerJ* **2018**, *6*, e4437. [CrossRef]
87. Goold, B.J. Open to all? Regulating open street CCTV and the case for "symmetrical surveillance". *Crim. Justice Ethics* **2006**, *25*, 3–17. [CrossRef]
88. Welsh, B.C.; Farrington, D.P. Public area CCTV and crime prevention: An updated systematic review and meta-analysis. *Justice Q.* **2009**, *26*, 716–745. [CrossRef]
89. Kleinberg, K.F.; Vanezis, P.; Burton, A.M. Failure of anthropometry as a facial identification technique using high-quality photographs. *J. Forensic Sci.* **2007**, *52*, 779–783. [CrossRef] [PubMed]
90. Keval, H.U.; Sasse, M.A. Can we ID from CCTV? Image quality in digital CCTV and face identification performance. *Mob. Multimedia Image Process. Secur. Appl.* **2008**, *6982*, 69820. [CrossRef]
91. Smith, R.A.; MacLennan-Brown, K.; Tighe, J.F.; Cohen, N.; Triantaphillidou, S.; MacDonald, L.W. Colour analysis and verification of CCTV images under different lighting conditions. *Image Qual. Syst. Perform. V* **2008**, *6808*, 68080. [CrossRef]
92. Bindemann, M.; Attard, J.; Leach, A.; Johnston, R.A. The effect of image pixelation on unfamiliar-face matching. *Appl. Cogn. Psychol.* **2013**, *27*, 707–717. [CrossRef]
93. Burton, A.M.; Wilson, S.; Cowan, M.; Bruce, V. Face recognition in poor-quality video: Evidence from security surveillance. *Psychol. Sci.* **1999**, *10*, 243–248. [CrossRef]
94. Ritchie, K.L.; White, D.; Kramer, R.S.S.; Noyes, E.; Jenkins, R.; Burton, A.M. Enhancing CCTV: Averages improve face identification from poor-quality images. *Appl. Cogn. Psychol.* **2018**, *32*, 671–680. [CrossRef]
95. Fysh, M.C.; Bindemann, M. The Kent Face Matching Test. *Br. J. Psychol.* **2018**, *109*, 219–231. [CrossRef]
96. Kramer, R.S.S.; Mohamed, S.; Hardy, S.C. Unfamiliar Face Matching With Driving Licence and Passport Photographs. *Perception* **2019**, *48*, 175–184. [CrossRef] [PubMed]
97. Burton, A.M.; White, D.; McNeill, A. The Glasgow Face Matching Test. *Behav. Res. Methods* **2010**, *42*, 286–291. [CrossRef]
98. Bindemann, M.; Attard, J.; Johnston, R.A. Perceived ability and actual recognition accuracy for unfamiliar and famous faces. *Cogent Psychol.* **2014**, *1*, 986903. [CrossRef]
99. Damjanovski, V. *CCTV from Light to Pixels*, 3rd ed.; Elsevier: Oxford, UK, 2014; ISBN 9780124045576.
100. Ward, D. Testing Camera Height vs. Image Quality; Pennsylvania, USA. 2013. Available online: https://ipvm.com/reports/testing-camera-height (accessed on 25 November 2021).
101. Mileva, M.; Burton, A.M. Face search in CCTV surveillance. *Cogn. Res. Princ. Implic.* **2019**, *4*, 1–21. [CrossRef]
102. Abudarham, N.; Yovel, G. Reverse engineering the face space: Discovering the critical features for face identification. *J. Vis.* **2016**, *16*, 40. [CrossRef] [PubMed]
103. Stephan, C.N. Perspective distortion in craniofacial superimposition: Logarithmic decay curves mapped mathematically and by practical experiment. *Forensic Sci. Int.* **2015**, *257*, 520.e1–520.e8. [CrossRef]
104. Stephan, C.N.; Armstrong, B. Scientific estimation of the subject-to-camera distance from facial photographs for craniofacial superimposition. *Forensic Sci. Int. Rep.* **2021**, *4*, 100238. [CrossRef]
105. British Security Industry. Video Surveillance Systems for Use in Security Applications: System Requirements—General. 2014. Available online: https://www.bsia.co.uk/publications/video-surveillance/ (accessed on 30 October 2021).
106. Cohen, N.; Gattuso, J.; MacLennan-Brown, K. *CCTV Operational Requirements Manual*; Cohen, N., Gattuso, J., MacLennan-Brown, K., Eds.; Home Office Scientific Development Branch: Sandridge, UK, 2009; ISBN 9781847269027.

107. Surette, R. The thinking eye: Pros and cons of second generation CCTV surveillance systems. *Policing* **2005**, *28*, 152–173. [CrossRef]
108. Kruegle, H. *CCTV Surveillance: Analog and Digital Video Practices and Technology*, 2nd ed.; Elsevier Butterworth–Heinemann: Burlington, MA, USA, 2007; ISBN 9780750677684.
109. Blunden, B. Anti-Forensics: The Rootkit Connection. In Proceedings of the Black Hat USA 2009, Las Vegas, NV, USA, 25–30 July 2009; pp. 1–44.
110. D'Orazio, C.; Ariffin, A.; Choo, K.-K.R. IOS Anti-Forensics: How Can We Securely Conceal, Delete and Insert Data? In Proceedings of the 47th Hawaii International Conference on System Sciences, Waikoloa, HI, USA, 13 October 2013; pp. 6–9. Available online: https://ssrn.com/abstract=2339819 (accessed on 30 October 2021).
111. Kissel, R.; Regenscheid, A.; Scholl, M.; Stine, K. *Guidelines for Media Sanitization*; US Department of Commerce, National Institute of Standards and Technology: Gaithersburg, MD, USA, 2014; Volume 800.
112. Ariffin, A.; Choo, K.K.; Yunos, Z. Forensic readiness: A case study on digital CCTV systems antiforensics. In *Contemporary Digital Forensic Investigations of Cloud and Mobile Applications*; Choo, K.-K.R., Dehghantanha, A., Eds.; Syngress: London, UK, 2017; pp. 147–162.
113. Gibelli, D.; Obertová, Z.; Ritz-Timme, S.; Gabriel, P.; Arent, T.; Ratnayake, M.; De Angelis, D.; Cattaneo, C. The identification of living persons on images: A literature review. *Leg. Med.* **2016**, *19*, 52–60. [CrossRef]
114. ENFSI. *Best Practice Manual for Facial Image Comparison*; ENFSI: Wiesbaden, Germany, 2018; Volume 1, Available online: https://enfsi.eu/wp-content/uploads/2017/06/ENFSI-BPM-DI-01.pdf (accessed on 30 October 2021).
115. Smith, S. CCTV Market Outlook 2017. *Cision PR Newswire*, 15 May 2014.
116. Wood, L. CCTV Cameras—Worldwide Market Outlook Report 2018–2026: Dome Cameras Dominate. *Businesswire* **2018**. Available online: https://www.businesswire.com/news/home/20180913005519/en/CCTV-Cameras---Worldwide-Market-Outlook-Report-2018-2026-Dome-Cameras-Dominate---ResearchAndMarkets.com (accessed on 30 October 2021).
117. Yip, A.W.; Sinha, P. Contribution of color to face recognition. *Perception* **2002**, *31*, 995–1003. [CrossRef]
118. Bindemann, M.; Burton, A.M. The role of color in human face detection. *Cogn. Sci.* **2009**, *33*, 1144–1156. [CrossRef] [PubMed]
119. Hunt, R.W.G. *The Reproduction of Colour*, 5th ed.; Fountain Press: London, UK, 1995.
120. MacDonald, L.W. Fatal Flaws: Uncertainty in the Interpretation of Colour in CCTV Images. *Ann. Br. Mach. Vis. Assoc.* **2007**, *7*, 1–11.
121. Bachmann, T. Identification of spatially quantised tachistoscopic images of faces: How many pixels does it take to carry identity? *Eur. J. Cogn. Psychol.* **1991**, *3*, 87–103. [CrossRef]
122. Bruce, V.; Henderson, Z.; Newman, C.; Burton, A.M. Matching identities of familiar and unfamiliar faces caught on CCTV images. *J. Exp. Psychol. Appl.* **2001**, *7*, 207–218. [CrossRef]
123. Costen, N.P.; Parker, D.M.; Craw, I. Spatial Content and Spatial Quantisation Effects in Face Recognition. *Perception* **1994**, *23*, 129–146. [CrossRef] [PubMed]
124. Costen, N.P.; Parker, D.M.; Craw, I. Effects of high-pass and low-pass spatial filtering on face identification. *Percept. Psychophys.* **1996**, *58*, 602–612. [CrossRef] [PubMed]
125. Viték, S.; Klíma, M.; Krasula, L. Video compression technique impact on efficiency of person identification in CCTV systems. In Proceedings of the Proceedings—International Carnahan Conference on Security Technology, Rome, Italy, 13–16 October 2014.
126. Utochkin, S. The Principles of CCTV Design in VideoCAD Interframe Compression. 2011. Available online: https://cctvcad.com/Files/the_principles_of_cctv_design_in_videocad_part5.pdf (accessed on 30 October 2021).
127. Stephan, C.N. Estimating the Skull-to-Camera Distance from Facial Photographs for Craniofacial Superimposition. *J. Forensic Sci.* **2017**, *62*, 850–860. [CrossRef]
128. Klima, M.; Fliegel, K. Image compression techniques in the field of security technology: Examples and discussion. In Proceedings of the 38th Annual 2004 International Carnahan Conference on Security Technology, Albuquerque, NM, USA, 11–14 October 2004; IEEE: Albuquerque, NM, USA, 2014; pp. 278–284.
129. Qi, X.; Liu, C. Mitigate compression artifacts for face in video recognition. In Proceedings of the Disruptive Technologies in Information Sciences IV, online. 27 April–8 May 2020; Blowers, M., Hall, R.D., Dasari, V.R., Eds.; SPIE: Bellingham, WA, USA, 2020; p. 25.
130. Ghazali, N.N.A.N.; Zamani, N.A.; Abdullah, S.N.H.S.; Jameson, J. Super resolution combiantion methods for CCTV forensic interpretation. In Proceedings of the 12th International Conference on Intelligent Systems Design and Applications, Kochi, India, 27–29 November 2012; pp. 853–858.
131. Stevenage, S.V.; Pitfield, C. Fact or friction: Examination of the transparency, reliability and sufficiency of the ACE-V method of fingerprint analysis. *Forensic Sci. Int.* **2016**, *267*, 145–156. [CrossRef]
132. Norell, K.; Läthén, K.B.; Bergström, P.; Rice, A.; Natu, V.; O'Toole, A. The Effect of Image Quality and Forensic Expertise in Facial Image Comparisons. *J. Forensic Sci.* **2015**, *60*, 331–340. [CrossRef] [PubMed]
133. FISWG. *Guide for Role-Based Training in Facial Comparison*; FISWG: Washington, DC, USA, 2020; Available online: https://www.fiswg.org/fiswg_guide_for_role-based_training_in_facial_comparison_v1.0_20200717.pdf (accessed on 30 October 2021).
134. Towler, A.; Kemp, R.I.; Mike Burton, A.; Dunn, J.D.; Wayne, T.; Moreton, R.; White, D. Do professional facial image comparison training courses work? *PLoS ONE* **2019**, *14*, e0211037. [CrossRef] [PubMed]

135. Davis, J.P.; Tamonyte, D. Masters of disguise: Super-recognisers' superior memory for concealed unfamiliar faces. In Proceedings of the 2017 Seventh International Conference on Emerging Security Technologies (EST), Canterbury, UK, 6–8 September 2017; IEEE: Canterbury, UK, 2017; pp. 44–49.
136. Russell, R.; Duchaine, B.; Nakayama, K. Super-recognizers: People with extraordinary face recognition ability. *Psychon. Bull. Rev.* **2009**, *16*, 252–257. [CrossRef] [PubMed]