Effects of Postmortem Decomposition on Face Recognition

David C. Cornett III, David S. Bolme
Oak Ridge National Laboratory
1 Bethel Valley Road, Oak Ridge, Tennessee 37831
cornettdciii@ornl.gov, bolmeds@ornl.gov

Dawnie W. Steadman
University of Tennessee, Knoxville, Department of Anthropology
1924 Alcoa Hwy, Knoxville, TN 37920
osteo@utk.edu

Kelly A. Sauerwein
Standards Coordinating Body
100 Bureau Drive, MS 8543, Bldg 224, A105, Gaithersburg, MD 20899
kelly.sauerwein@nist.gov

Tiffany B. Saul
Middle Tennessee State University, Forensic Institute for Research and Education
1301 East Main St., Box 89, Murfreesboro, TN 37132
tiffany.saul@mtsu.edu

Abstract

Although viable tools are available for the identification of unknown deceased individuals, recognition rates with these methods are greatly impacted by the degree to which decomposition has occurred. Therefore, identifying highly decomposed remains poses a major challenge. This paper analyzes the effect of facial decomposition on the recognition rates of several facial recognition commercial-off-the-shelf systems and research-grade systems, as well as algorithms contained in a custom recognition library. The custom dataset of facial images used in the experiment is composed of 42 subjects at stages of decomposition ranging from recently deceased to later stages where the soft tissues are severely decomposed and facial features are deformed. It is shown that an algorithm's ability to correctly detect a decomposing face is a crucial first step that not all face models can accurately handle. However, some of the evaluated Convolution Neural Network (CNN)–inspired methods provide promising results even in cases of severely decomposed faces.

1. Introduction

Identification of human remains is a key challenge in forensics science. This paper presents an evaluation of multiple face recognition algorithms used to identify decomposing bodies. Due to the ease with which facial photographs can be made available and the improved accuracy of recognition algorithms, facial recognition has the potential to accelerate investigations by instantly identifying remains. In this study we focus on the utility of face recognition algorithms by defining a set of postmortem-specific quality categories and evaluating the accuracy of face recognition algorithms. While recently deceased individuals with no facial damage can be easily recognized with modern algorithms, this study focuses on significantly decomposed faces up to the point where facial features are completely "deformed"

This manuscript has been authored in part by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

and other recognition techniques will need to be applied.

Deep learning methods have reshaped the research landscape of face recognition in almost all aspects such as algorithm designs, training/test datasets, application scenarios, and even the evaluation protocols [1]. These methods have performed well despite such challenges as varying poses, inconsistent illumination, and other "in the wild" issues. However, no known research has evaluated the performance of deep neural networks when applied to the drastic facial changes that occur during decomposition. Such identification is of major importance to forensic pathologists when attempting to identify an unknown deceased individual.

This experiment uses a custom dataset of subject faces in varying degrees of decomposition. Facial images of a subject were collected upon intake to the facility (typically at 1 to 2 days after death) and then daily throughout the course of decomposition. This collection process is structured such that the effects of decomposition are observed in a gradual manner.

These images were then subdivided into six subjective levels of decomposition quality to ascertain at what point a decomposing face can no longer be detected or accurately identified. Numerous commercial-off-the-shelf (COTS), government-off-the-shelf (GOTS), research, and custom algorithms were evaluated as to their ability to both correctly detect decomposing faces and identify them as being the intake and high-level images (ground truth) of the same subject.

2. Background

Postmortem identification of humans is of utmost importance to law enforcement agencies. The effects of decomposition greatly hamper the process of accurately matching an unknown individual to someone in a database of missing persons. Comparisons of antemortem and postmortem data of fingerprints, DNA, and dental records (primary means of identification) are considered to be the most reliable methods of identifying an individual [2]. Although fingerprints and DNA have become some of the best tools used in forensic investigations, they are not always helpful due to either contamination, damage to the body, or a lack of these biometrics from living subjects. Additionally, these methods are only utilized by trained experts, whereas faces provide a reasonable identification method for non-experts, provided decomposition is not too severe. Therefore, a system that could perform facial identification of an unknown deceased individual would be a valuable tool for both skilled forensic technicians and non-experts.

The dataset used in this paper comes from an ongoing study by researchers at the University of Tennessee. Fingerprints, irises, and facial images are collected on a daily basis from decomposing individuals as described by Sauerwein *et al*. [3]. This research has presented findings related to the accuracy of biometric identification on this dataset [4, 5]. The study presented in this paper is focused exclusively on facial biometrics and includes more donors and additional recognition algorithms (particularly deep learning–based methods) and is therefore an important follow-on to the previous research.

Much of the postmortem research in facial identification has used the methods of face reconstruction and face superimposition starting with a clean human skull. Aulsebrook *et al*. [6] provides an in-depth survey of research covering these two techniques.

Face reconstruction may be in a two-dimensional form (sketches) or three-dimensional assessment with 3D computational models or physical sculptures. Reconstruction relies on the morphological features of the skull coupled with knowledge of how the facial soft tissues are expressed around these features. This method therefore requires both the expertise of a forensic scientist coupled with the expertise of the artist who reconstructs the face. Although face reconstruction can produce very realistic results, there is no standardized method, so several decisions are made during the reconstruction that can lead to unreliable results for identification purposes.

Two-dimensional superimposition techniques are considered to be reconstructions in that they attempt to supply a face for an unknown skull [6]. This method requires that the suspected true identity (antemortem image) of the subject is known so that it can be superimposed onto the skull for comparison analysis. Therefore, superimposition is effectively a 1:1 biometric where the researcher's goal is to verify a suspected identity as opposed to identifying the unknown remains against a database of possible subjects.

Although both of these methods provide effective tools to assist in postmortem identification, our research departs in three primary ways: (1) both methods start from skulls, whereas our research focuses on faces where at least some soft tissue is still present, (2) face reconstruction is only used to garner leads in a case, not for verification and identification, and (3) superimposition is only used as an exclusionary tool (not for verification). Our research is aimed specifically toward biometric identification with automated methods.

The iris of the eye has been shown to be highly accurate in typical biometric identification systems. Previous work [4] has suggested that the rate of decomposition of the iris can be highly variable given the environment in which the body is found. The iris is generally thought to be viable for only a few days after death. Other work [3] related to decomposing iris identification has demonstrated that the iris can be viable for 2–34 days. Recent work [7] has shown that irises kept in mortuary conditions remain viable for identification significantly longer than those subjected to various weather conditions. Although this paper is not focused on identification via irises, the work in [7] also indicates the potential for applying deep learning methods to decomposing biometrics modalities.

3. Dataset

The 42 human subjects included in this study are whole body donors to the body donation program at the Forensic Anthropology Center (FAC) of the University of Tennessee, or "Body Farm," a nearly 3-acre natural outdoor laboratory to study human decomposition. Donors either self-donate by registering with the FAC while living or are donated by families at or around the time of death. All donors to the Body Donation Program at the FAC have signed (or a family member has signed) forms that provide permission for research. Given the sensitivity of recognizable faces in this project, it is notable that we have not included any photos in the paper, nor will recognizable photos be shown at the conference presentation out of respect for the privacy of the donors. Moreover, the research project, for which this paper is a subset, has been reviewed by the University of Tennessee Internal Review Board.

Donors are received at intake where biometrics and other sampling procedures are carried out and then placed at the FAC. Each donor for this study was placed supine and biometrics (iris, fingerprint and face) were taken daily until decomposition precluded image capture.

3.1. Facial Changes during Decomposition Affecting Recognition

The set of visible changes that occur during decomposition span numerous variations. Typically a relaxed expression and discoloration is the first change. The skin loosens and the eyes and mouth become slack and produce expressions not commonly displayed while living. Severe discoloration begins, and the skin can begin to slip off. Next, is bloating and, depending on the season, insect infestation begins. Scavengers may also remove the eyes and other portions of the face. Finally, most of the recognizable facial features have either slipped into a unnatural position or have been removed, and the face can be completely occluded by insect infestation.

Expression changes are one of the earliest noticeable variations about the face. Intake photos are often taken 2–3 days after death and are usually post-rigor when muscles are relaxed. This results in a blank expression where the mouth often drops open, typically not showing teeth. The experimental protocol requires the eyes to be open during the photographs, and in some cases they are held open using wire speculum. The focus of the eyes' gaze tends to be centered but the eyes can sometimes rotate in the sockets and in some instances may point in different directions. Typically the expression persists through later stages of decomposition.

Discoloration happens soon after death as the blood stops circulating and drains to the lower parts of the body. In this dataset all bodies are supine; thus, the blood pools away from the face. The paleness is less visible in the indoor photos but becomes obvious in the outdoor natural light, which causes the faces to appear gray. This could affect algorithms that rely on skin tone for detection or recognition.

The skin often darkens when it begins to break down, similar to bruising, and may have a marbled appearance due to the blood vessels in the face. This often progresses in cooler environments to mummification where the skin dehydrates and becomes a lighter shade of gray. In some cases where decomposition is faster, the skin can become very dark.

Bloating happens when bacteria in the body release gas but did not often affect the faces seen in this dataset. However when present, bloating seems to mostly affect the jaw and cheeks.

Insects, especially fly maggots, appear in the facial orifices after 3–5 days in the warmer months. Eye colonization often appears first, and eyes will no longer be usable features a day after colonization begins. The insects appear as pale masses in the orifices and expand later to cover the skin of the face. The insects consume soft tissues quickly, and the viability of standard face recognition approaches will deteriorate rapidly.

Tissue Damage can occur early on when the facial features are damaged either by trauma at the time of death or by scavengers that are attracted to the bodies early in the decomposition process. Scavengers often target the face and hands, which causes problems for identification [8]. As the decomposition process proceeds, the skin starts to break down. There are often rips, separation, and slippage in the skin. The eyes also dehydrate, are scavenged, or will sink into the skull, causing additional challenges to recognition. Facial features become distorted, and bone can be exposed as the soft tissues break down.

3.2. Subjective Quality Assessment and Dataset Organization

Decomposition rates are highly affected by such factors as weather patterns, temperature, humidity, insect infestation, etc. Therefore, to determine the effects of decomposition upon facial recognition, the subject images were manually sorted into decomposition quality sets. These sets represent subjective levels of decomposition organized at a meta level so that recognition rates could be evaluated without including subject exposure parameters. These subject quality levels were intake, high, medium, low, poor, and deformed.

- **Intake** images are those that are captured when the subject arrives at the receiving facility and provide the closest likeness to an antemortem image.

- **High**-quality images are are captured after the first day or two of outdoor exposure when minimal decomposition has begun.
- **Medium**-quality images show moderate decomposition, which occurs sooner in high-temperature conditions than in winter or fall.
- **Low**-quality images are those in which all facial features are still intact regardless of the severe state of decomposition.
- **Poor**-quality images are those in which facial features are missing or structural damage has occurred and also includes partial occlusion from insect activity.
- **Deformed**-quality images are those with extremely damaged or missing facial features and in which high occlusion from insects and putrefaction has occurred.

Figure 1 shows the distribution of subject decomposition quality levels across the whole data set. As shown, the number of images at each subjective quality level are relatively equal.

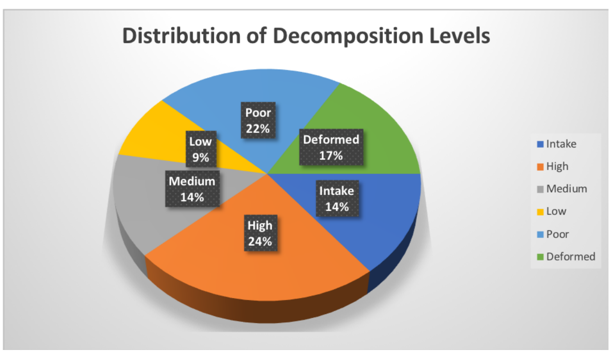

Figure 1. Percentages of the dataset across different subject quality levels.

4. Experimental Setup and Results

Currently, the dataset is composed of 42 subjects and includes a total of 544 images and is collected as described in [3]. The subjects were placed in an outdoor facility over the course of a year to account for differences in decomposition due to temperature and weather patterns. Because the subjects were exposed to a host of different parameters, image quantity and quality varied even for the same subject. The distribution of quality for the individual images per subject is shown in Figure 2.

Three COTS systems and three GOTS systems were also tested as part of this study. Of these, GOTS 1 uses a proprietary algorithm and predates the deep learning revolution. GOTS 2 and 3 are recent algorithms known to leverage deep learning for detection and recognition. COTS 2, 3, and 4 are recently released algorithms that use an unknown proprietary technology.

Due to agreements and licenses for both the COTS and GOTS systems, the underlying algorithms must be kept anonymous and therefore we are limited to the degree of technical detail of these systems that we can provide in this paper. The details of these systems can be made available to government and law enforcement agencies with appropriate approval.

For evaluation we have additionally used two deep learning face recognition algorithms based on open source technology. One algorithm comes from *dlib*[9]. The other algorithm is a combination of the deep learning face detector from *dlib* and the VGG ResNet–based facial feature descriptor from [10]. Because these libraries just provide parts of a face recognition solution, the FaRO library [11] was used to connect the components into a single complete system and standardize those algorithms as a reproducible baseline. The FaRO library is an open-source face recognition project which is designed to provide a full face recognition pipeline (detect, extract, match) where each component of this pipeline can be combined with the associated utility from open source projects.

4.1. Face Detection Analysis

The first challenge in automatic face recognition is detection. During the early stages of decomposition, the main features of the face remain unchanged. The faces classified as intake or high quality in the early stages seem to still have very high detection rates. In these stages, relaxation of muscles and pallor have little effect on detection. Certainly, as the features of the face enter the later stages of decomposition and become deformed, the features defining the face are missing or distorted. The detection of faces in these stages of decomposition can be difficult for even the best algorithm. It is likely even humans will fail to find the face if not provided the right context.

For the purposes of this study, face detection was run in a “best” mode, which assumes one face was in each image and one detection rectangle and template was always returned from each image. This ensures that each image can always be compared with others. In some cases, the bounding boxes returned by the detectors were wrong and these were scored by hand to estimate detection accuracy rates. Using the “best” detection mode also allowed the generation of match scores at the end of the process; however, it is expected that most images with inaccurate detection rectangles will match poorly.

Figure 3 shows the results of eight face recognition systems compared with the total number of images in each quality category. Here the detection algorithms were set to a “best” mode where the detection with the highest score was used. In cases where no detections were returned from

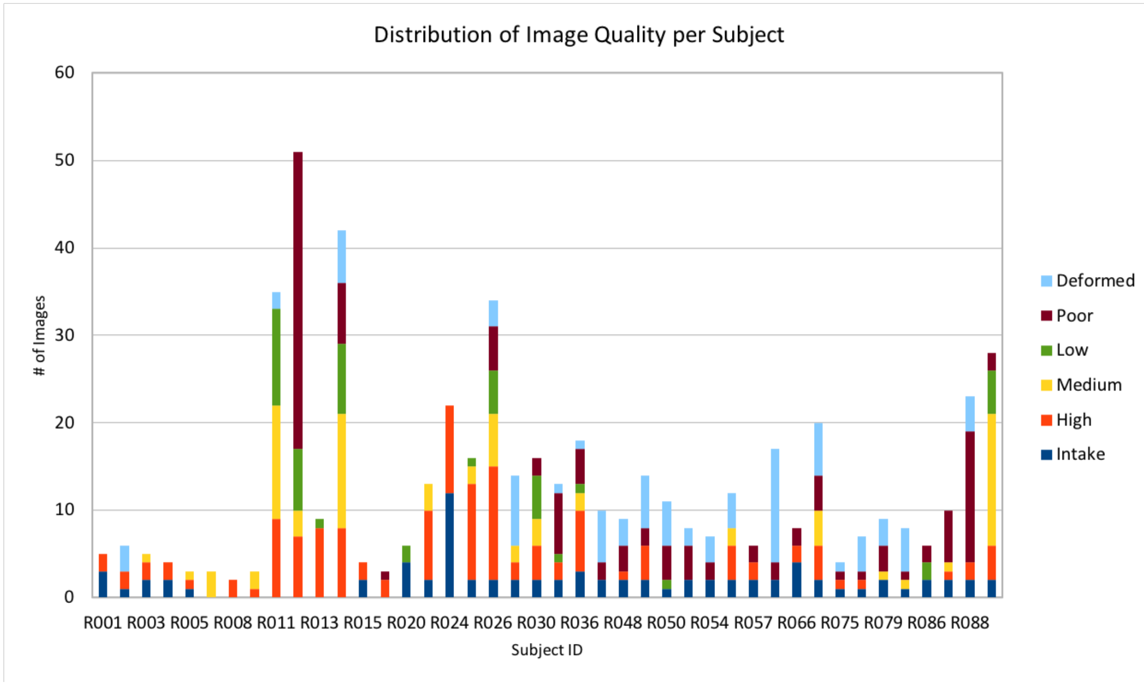

Figure 2. Distribution of images per subject.

the algorithm, a heuristic was used to determine a bounding box centered in the image that would roughly select the face if the image were a passport photo.

The older algorithm (GOTS 1) performs poorly even on intake and high-quality faces, while the newer and deep learning–based algorithms perform well on intake, high-, medium-, and even low-quality faces. Four algorithms lose accuracy starting with poor-quality faces. All algorithms perform poorly on deformed faces. GOTS 3 however performs the best across the board and has a significant advantage, especially for poor- or deformed-quality faces.

An additional parameter affecting the detection rates is the detection threshold used by the algorithms. In our study we examined the effect of using two thresholds: the algorithm's default threshold (an algorithm may not detect a face in the image) and "best" mode, where the algorithm always makes one estimate at face detection. This comparison of detection thresholds was needed to standardize the evaluation across the algorithms given that detection rates could vary widely from algorithm to algorithm due to the decomposition of facial features. In the best mode plot, all the calculated bounding boxes were manually verified to correct facial detection. Figure 4 explores the detection rates between three of the tested algorithms and the effects of threshold value in accurate facial detection. Note that the GOTS 3 algorithm when run in default mode found multiple false detections in many of the images.

4.2. Face Identification Analysis

Automatic identification is analyzed in this section. For these experiments, images that were captured at intake or fell into the high categories are used as a gallery. Images in the medium, low, poor, and deformed categories are used as probes. We are most interested in the small false accept rates. Considering real world applications, low false accept rates in the ranges of 0.001 to 0.01 are most relevant.

Figure 5 shows four selected Receiver Operating Characteristic (ROC) curves from the algorithms. The plots show significantly different performance among the algorithms. The COTS 3, FARO-Dlib VGG2, and GOTS3 algorithms performed well on the recognition task when matching the medium- and low-quality images, but performance dropped quickly for the poor- and deformed-quality images. COTS 3 and GOTS 3 show what we believe are very good results on this datasets; however, there are notable differences. GOTS 3 produced the best results on medium- and high-quality data, but COTS 3 performance with the poor-quality images was better, and COTS 3 seemed to produce flatter ROC curves at low false acceptance rate (FAR). Interestingly, GOTS 3 also appeared to perform slightly better with the low-quality images over the medium-quality images at more challenging FAR levels.

All algorithms failed to recognize the deformed category of faces. This is not surprising given the extreme challenges posed in this portion of the dataset. At some point in the decomposition process, automatic face recognition techniques

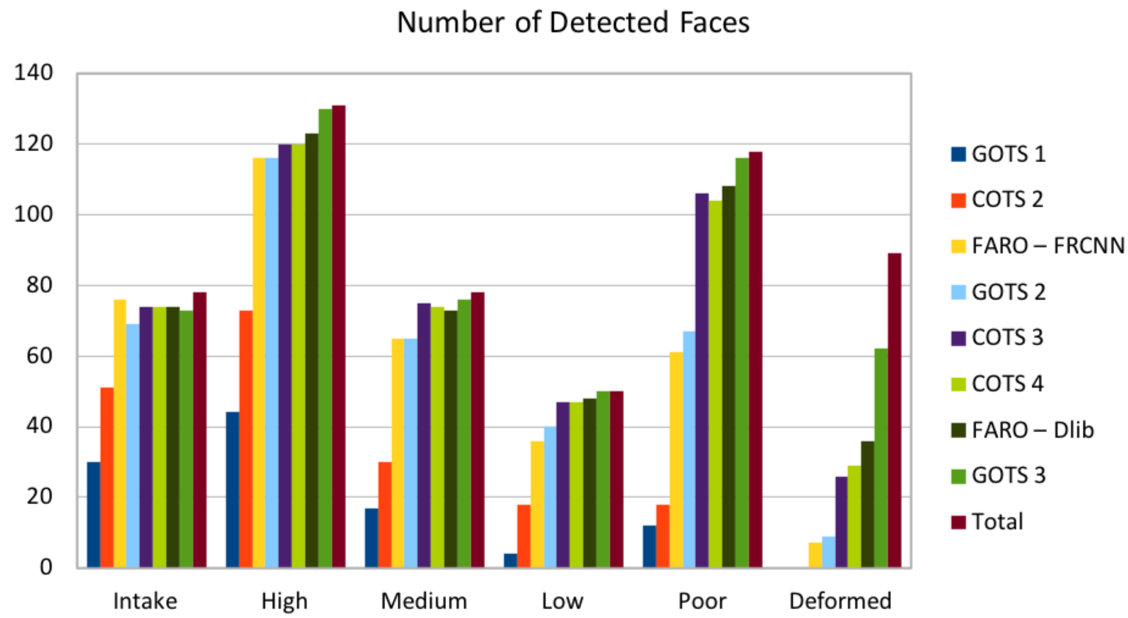

Figure 3. Detection results from multiple algorithms.

will fail. It appears that the criteria selected for the deformed category well describe the point of failure for these algorithms.

One behavior that was observed with the FARO-Dlib VGG2 algorithm was how the matching threshold relates to the false accept rate. The thresholds were calibrated using the Labeled Faces in the Wild (LFW) dataset [12] but were significantly different on the harder decomposing face images. The algorithm thus produced much higher-than-expected false accept rates. When the images were examined, falsely matched pairs seemed to show similar discoloration patterns and expressions, which indicates those features may be seen by the algorithm as indicators of identity and not of decomposition. This was not evaluated on other algorithms and is worth further study. As a result, thresholds should be checked and adjusted when used with postmortem images.

5. Conclusion

Although identification of decomposing faces poses a significant challenge, particularly at later stages of decomposition, the use of deep neural networks has been demonstrated in classical face recognition algorithms. We believe that similar advances can be made for decomposing faces if sufficient datasets can be obtained for algorithm retraining. Four algorithms (DLIB, DLIB+VGG ResNet, COTS 3, and GOTS 3) have demonstrated performance on this dataset, suggesting good results in forensics applications. The results demonstrate that modern algorithms appear to detect correctly for decomposing faces even up to the category that we considered poor quality, although at that point recognition rates deteriorated.

This paper advances the related work in [4], [3], and [5] by the inclusion of recent improvements in facial recognition which leverage the recent advances in convolutional networks. In particular, this work includes the performance of COTS 3 and GOTS 3 algorithms, which are some of the top performers in other evaluations and are representative of the 2019 state of the art. These deep learning models show significant accuracy improvements over the previous related work. Additionally, this paper also presents performance curves for this dataset which allow this application to be compared with other evaluations using standard methods utilized in most academic biometric publications.

The proposed quality measures are, as they have been defined, predictive of face recognition performance as shown in this paper. However, there is still a need for more quantitative and automatic assessment of these issues. These quantitative measures have been difficult to define due to the limitations arising from the small number of subjects in the dataset. This current study contains 42 subjects, which limits the ability to statistically quantify facial changes. While expression and coloration changes are exhibited in most subjects, the type and appearance of change can vary significantly due to the environmental conditions surrounding the exposure of the various subjects. For example, insects are only present during warmer months and bloating only significantly affects a few of the subjects.

The quality labels proposed in this study correlate well with recognition accuracy, so a network is being trained to

Figure 4. Best detection mode vs. default threshold.

automatically categorize a subject image into one of the six quality labels. This will serve to categorize new subjects that are added to the dataset as well as mitigate the subjective aspect from the human rating of quality.

6. Acknowledgments

This research was supported in part by an appointment to the Oak Ridge National Laboratory Post-Bachelor's Research Associate Program, sponsored by the US Department of Energy and administered by the Oak Ridge Institute for Science and Education.

References

[1] Weihong Deng Mei Wang. Deep face recognition: A survey. *arXiv preprint arXiv:1804.06655*, 2019.

[2] Zuzana Caplova, Zuzana Obertova, Daniele M Gibelli, Danilo De Angelis, Debora Mazzarelli, Chiarella Sforza, and Cristina Cattaneo. Personal identification of deceased persons: An overview of the current methods based on physical appearance. *Journal of forensic sciences*, 63(3):662–671, 2018.

[3] Kelly Sauerwein, Tiffany B Saul, Dawnie Wolfe Steadman, and Chris B Boehnen. The effect of decomposition on the efficacy of biometrics for positive identification. *Journal of forensic sciences*, 62(6):1599–1602, 2017.

[4] D. S. Bolme, R. A. Tokola, C. B. Boehnen, T. B. Saul, K. A. Sauerwein, and D. W. Steadman. Impact of environmental factors on biometric matching during human decomposition. In *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–8, Sep. 2016.

[5] Tiffany B. Saul, Kelly Sauerwein, Dawnie Wolfe Steadman, and Chris B. Boehnen. Postmortem identification from physiological biometrics: A study of fingerprints, irises, and facial images. Poster: American Academy of Forensic Sciences, 2016.

[6] WA Aulsebrook, MY İşcan, JH Slabbert, and

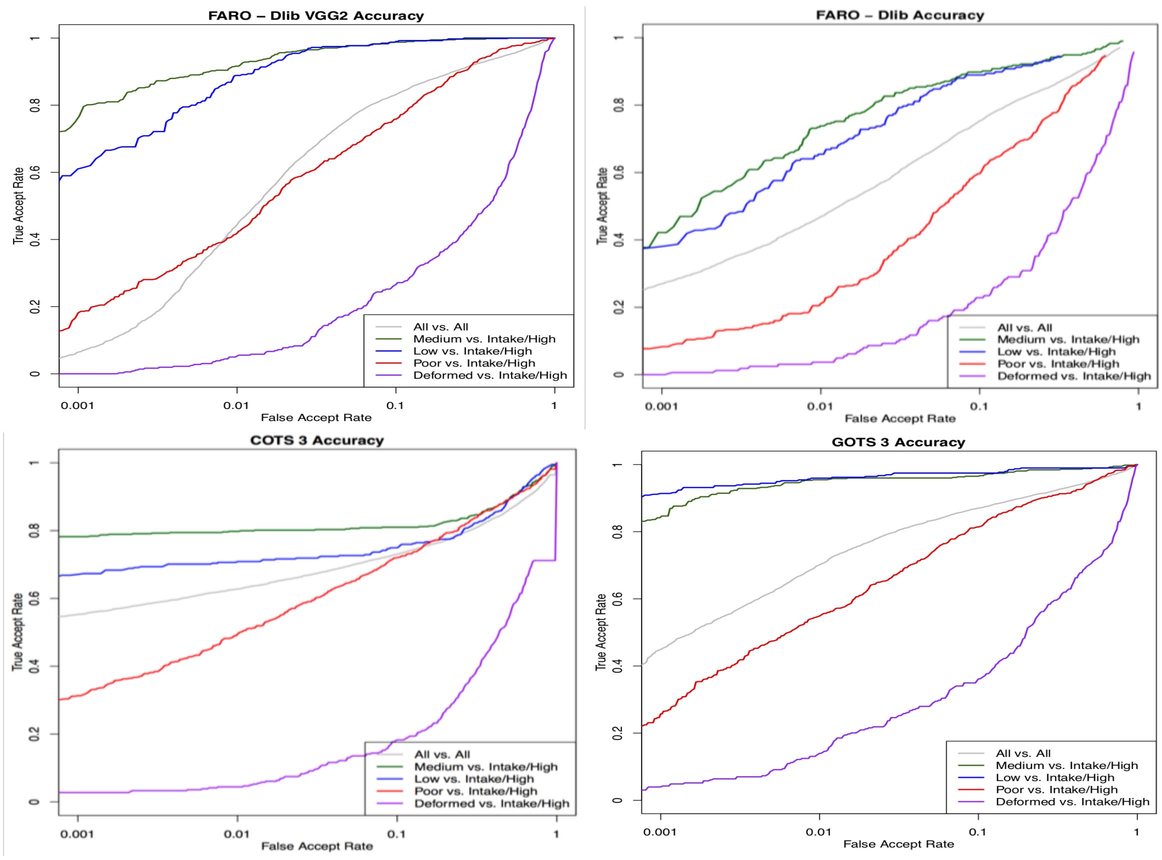

Figure 5. Selected algorithm ROCs. COTS 3 showed good identification performance even at low FAR for everything but deformed faces. FARO-Dlib VGG2 also performed well for a completely open source solution. GOTS 3 identification performance was the best for the medium- and low-quality faces.

P Becker. Superimposition and reconstruction in forensic facial identification: a survey. *Forensic science international*, 75(2-3):101–120, 1995.

[7] Mateusz Trokielewicz, Adam Czajka, and Piotr Maciejewicz. Iris recognition after death. *IEEE Transactions on Information Forensics and Security*, 14(6):1501–1514, 2019.

[8] Robert W Mann, William M Bass, and Lee Meadows. Time since death and decomposition of the human body: variables and observations in case and experimental field studies. *Journal of Forensic Science*, 35(1):103–111, 1990.

[9] The dlib c++ library. http://dlib.net/, 2019.

[10] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *International Conference on Automatic Face and Gesture Recognition*, 2018.

[11] David Bolme, Nisha Srinivas, and David Cornett III. Face Recognition from Oak Ridge (FaRO). https://github.com/ORNL/faro, 2019.

[12] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*, 2008.