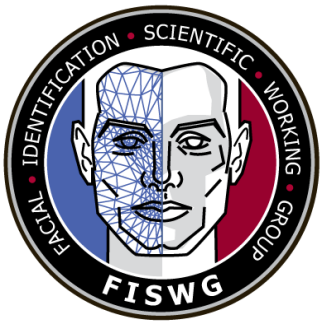

Disclaimer:

As a condition to the use of this document and the information contained herein, the Facial Identification Scientific Working Group (FISWG) requests notification by e-mail before or contemporaneously to the introduction of this document, or any portion thereof, as a marked exhibit offered for or moved into evidence in any judicial, administrative, legislative, or adjudicatory hearing or other proceeding (including discovery proceedings) in the United States or any foreign country. Such notification shall include: 1) the formal name of the proceeding, including docket number or similar identifier; 2) the name and location of the body conducting the hearing or proceeding; and 3) the name, mailing address (if available) and contact information of the party offering or moving the document into evidence. Subsequent to the use of this document in a formal proceeding, it is requested that FISWG be notified as to its use and the outcome of the proceeding. Notifications should be sent to: chair@fiswg.org

Redistribution Policy:

FISWG grants permission for redistribution and use of all publicly posted documents created by FISWG, provided that the following conditions are met:

Redistributions of documents, or parts of documents, must retain the FISWG cover page containing the disclaimer.

Neither the name of FISWG, nor the names of its contributors, may be used to endorse or promote products derived from its documents.

Any reference or quote from a FISWG document must include the version number (or creation date) of the document and mention if the document is in a draft status.

Facial Recognition Systems Operation Assurance: Manual Facial Localization

1. Scope

1.1 The scope of this document is to provide a detailed process with examples of testing and adjusting operational workflows for variations in facial image quality. Testing for and verifying scoring variations when image quality varies is critical so that proper facial search system workflows can be properly adjusted.

1.2 Being able to inspect images with low quality that could be improved through proper facial localization is the focus of this document. This document does not address any image processing but focuses solely on verifying facial localization was done correctly at enrollment.

1.3 Topics outside the scope of this document include, but are not necessarily limited to, system setup, system tuning, workflow management and improvement, and proof of concept pilots.

2. Referenced Documents

2.1 *ASTM Standards:* [1]

E2916 Terminology for Digital and Multimedia Evidence Examination

E2825 Standard Guide for Forensic Digital Image Processing

2.2 *Other Standards:*

ANSI/NIST- ITL-1-2011 Data Format for the Interchange of Fingerprint, Facial & Other Biometric Information[2]

[1] For referenced ASTM standards, visit www.nist.gov/osac/astm-launch-code, or the ASTM website, www.astm.org, or contact ASTM Customer Service at service@asstm.org. For Annual Book of ASTM Standards volume information, refer to the standard's Document Summary page on the ASTM website.

[2] Available from National Institute of Standards and Technology (NIST), 100 Bureau Dr., Stop 1070, Gaithersburg, MD 20899-1070, www.nist.gov/programs-projects/ansinist-itl-standard

3. Terminology

3.1 *Definitions:*

3.1.1 For terms relating to digital and multimedia evidence, refer to Digital and Multimedia Evidence Examination Terminology E2916.

3.2 *Definitions of Terms Specific to This Document:*

3.2.1 Doppelganger–a biologically unrelated look-alike.

3.3 *Acronyms:*

3.3.1 *FR*–Facial Recognition

3.3.2 *FRS*–Facial Recognition Systems

3.3.3 *CMC*–Cumulative Match Characteristic

3.3.4 *ROC*–Receiver Operating Characteristics

3.3.5 *DET*–Detection Error Tradeoff

3.3.6 *FAR*– (false acceptance rate)

3.3.7 *FMR*–False Match Rate proportion of the completed biometric non-mated comparison trials that result in a false match. This will be referred to as FAR and does not include errors from images which do not create valid templates.

3.3.8 *FNMR*–False Non-Match Rate proportion of the completed biometric mated comparison trials that result in a false non-match. This will be referred to as FRR (false reject rate) and does not include errors from images which do not create valid templates.

3.3.9 *IPD*–Inter-pupillary Distance

4. Summary of Guide

4.1 This document provides guidelines and techniques to help administrators of automated facial recognition systems (FRS) produce recognition statistics on the facial recognition systems, which can be used to improve overall biometric performance.

4.2 The intended audience of this document is system owners, system users, and system administrators of existing automated facial recognition systems.

4.3 This document is a continuation of these FISWG documents:

4.3.1 “Understanding and Testing for Facial Recognition Systems Operation Assurance”

4.3.2 “Facial Recognition Systems Operation Assurance: Part 2, Identity Ground Truth”

4.3.3 “Facial Recognition Systems Operation Assurance: Part 3, Image Quality Assessment”

4.4 The issues presented in this document form a base for other considerations and advanced topics when testing (e.g., system setup and tuning) which will be covered in future FISWG documents.

5. Significance and Use

5.1 Introduction

5.1.1 When doing accuracy assessments, a critical step is to understand how image quality affects the processing of the imagery to be enrolled or searched. Regardless of the specific facial biometric algorithm used, there will be some facial imagery which causes facial localization errors which then produces biometric templates of no operational value. Performing manual facial localization on images with improper localization can improve the biometric accuracy for images that need these manual operational steps. Enrolling facial images with known facial localization errors will only increase the risk of having false positive and false negative candidates in search results.

5.1.2 Most of the work in these processes is on creating the testing frameworks and understanding how to repeatedly run tests, make corrections, and conduct retesting with what has been learned. Once the frameworks and the processing are understood, then the agency can make diligent progress, but it takes time and focus. The outcomes are worth the time spent as you begin to understand how the data interacts with the algorithms which give the agency the ability to validate and trust the solution through rigorous testing and objective metrics.

5.1.3 Setting up frameworks to do enrollment and searching recording results is fairly mechanical as you learn the facial algorithms and the data sets to develop proper accuracy assessments. Understanding the data and building frameworks to analytically qualify the results is not trivial but must be done so effective operational metrics can be derived and applied.

5.2 Important Notes

5.2.1 Care should be taken in selecting data sets to assess. It is recommended to select data sets which:

- Have operational relevancy
- Have consistent image quality aspects: type of capture, size of images, subject poses, etc.
- Have sufficient identities and images to test with. This decision will be agency specific. This should include associated identity truth information.

5.2.2 The data set used for this document is the LFW (Labeled Faces in the Wild) data set available at: https://vis-www.cs.umass.edu/lfw/. See section “LFW Data Set Information “for more details in referenced document [2]. Conceptually any other facial data set with identity ground truth can be used.

5.3 LFW is a widely used open source data set which will work well for this specific document. Information on this data set includes:

- Smaller but consistent image sizes and file formats
- Over 5,700 identities and over 13,000 images
- Has a wide range of subjects: sex, pose, lighting, etc.
- Stated identity ground truth errors

6. Procedure

6.1 Manual Facial Localization Process

6.1.1 Ensure the data set has verified ground truth and the facial image quality assessment has been done so that an image quality range is known and can be used to separate images with low quality for inspection.

6.1.2 Using the facial image quality metrics derived from the images, determine a threshold of images which need to be inspected. The image quality metrics to use will be vendor dependent. Care should be taken to select the quality metrics which have the largest value in doing an image quality assessment. Consulting with the provider of the facial algorithm is recommended.

6.2 Identify the images to inspect. It is useful to mark all these images with the found eye locations so manual inspection can be performed.

6.3 Manually review the images and segment out the images with obvious facial localization errors that can be corrected with manual placement of the eyes. Certain images may not be applicable to this process based on the capabilities of the facial algorithm being used. An example here is +/- 90 degree profile images or high yaw images where both eyes are not present. Checking with the biometric vendor may assist in this specific step.

6.4 Determine the proper eye locations for the selected images.

6.5 Enroll the facial images into a facial search gallery.

6.6 Search the facial images against the facial gallery. The number of candidates returned for this document was 100. This number may vary with agency specifics. Do not use any scoring thresholds.

6.7 Search the facial images with the manual eye locations against the facial gallery. The number of candidates returned for this document was 100. This number may vary with agency specifics.

6.8 For each search group (i.e., original and manual eye locations) plot the biometric performance results and compare them. This document uses these plots:

- FAR: False Accept scoring
- FRR: False reject scoring
- DET: Detection Error Tradeoff
- CMC: Cumulative Match Curve

6.9 Process Outcomes

6.10 Step 2 Outputs: The image quality of the LFW data set was recorded from referenced document 3 (Image Quality Assessment).

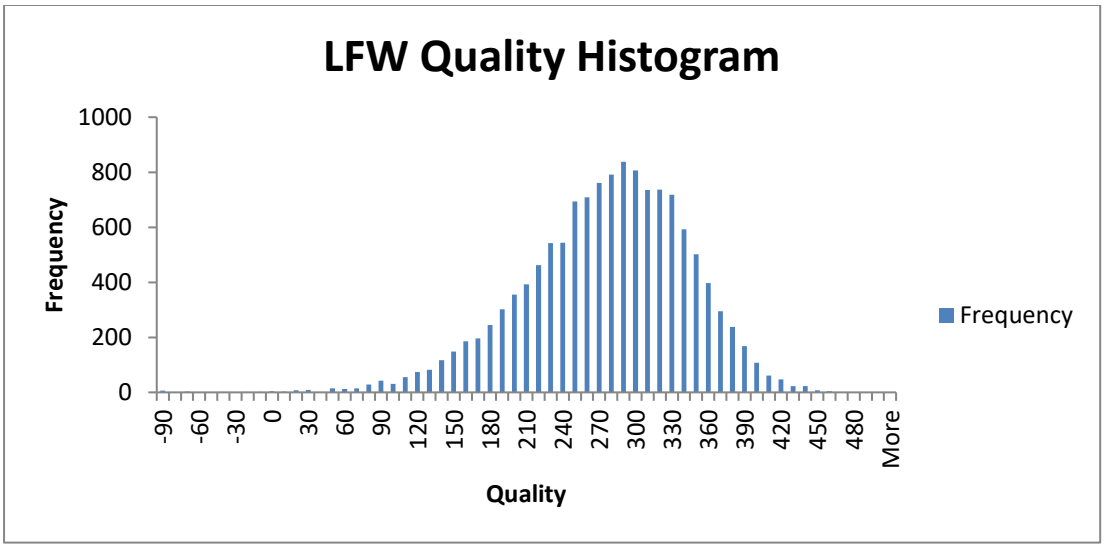

Figure 1: LFW Quality (linear Y axis)

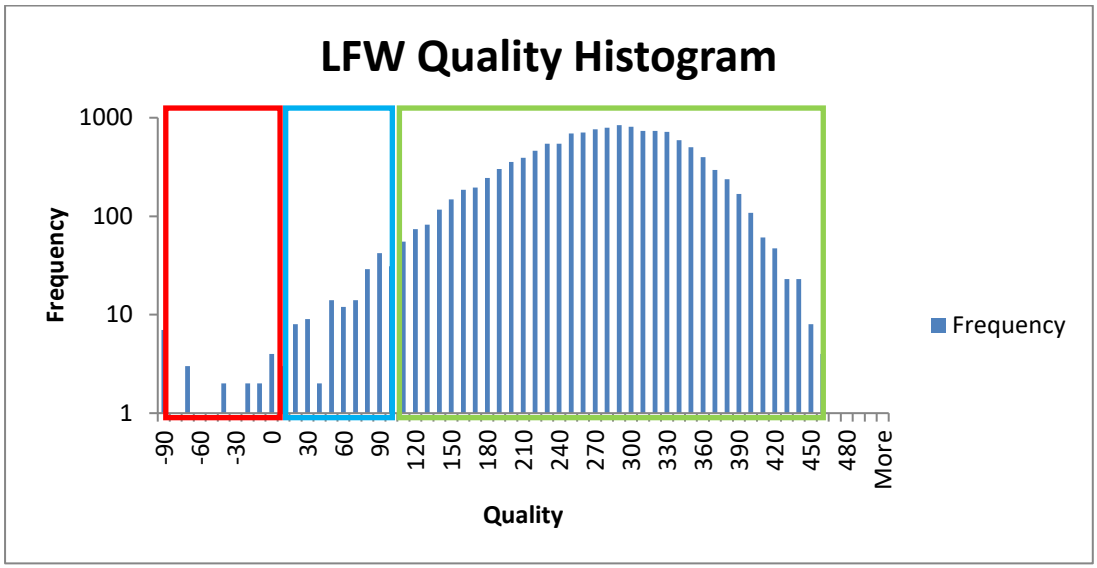

Figure 2: LFW Quality (logarithmic Y axis)

All LFW with a quality score less than 100 were selected producing 182 images.

- All images with a quality less than 0
- All images with a quality between 0 and 100
- All images with a quality greater than 100

6.11 Step 3 Outputs: The 182 images with an image quality of less than 100 were identified and copied into a unique directory.

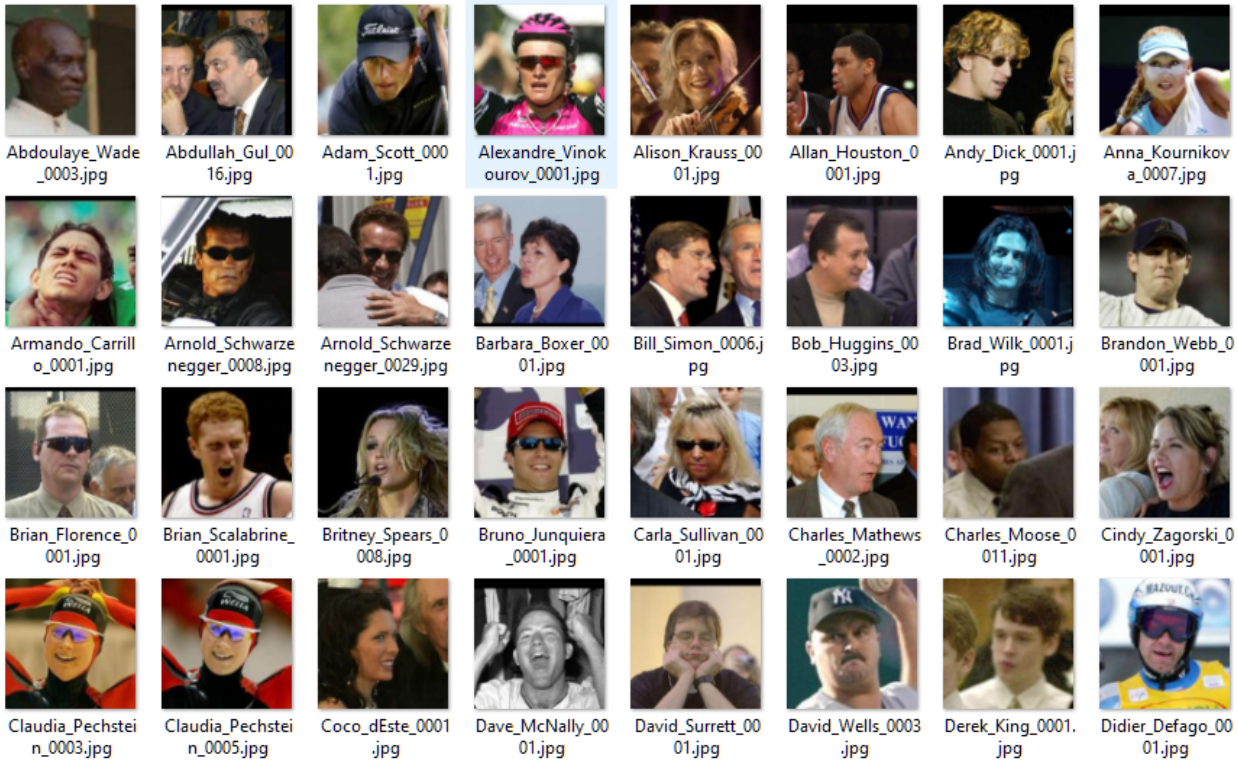

Figure 3: LFW Images below an image quality threshold

6.11.1 Step 4 Outputs: The following images show obvious facial localization errors which could be improved through manual alignment. From the 182 images, 32 images were believed to have potential using manual facial localization of the eyes.

Figure 4: LFW Images which may be improved though manual localization

The following images show low facial image quality which may not show improvement through manual alignment.

Figure 5: LFW Images which may not be improved though manual localization

6.11.2 Step 8 Outputs: Once the original and manually aligned images were searched, accuracy plots were created so they can be analyzed.

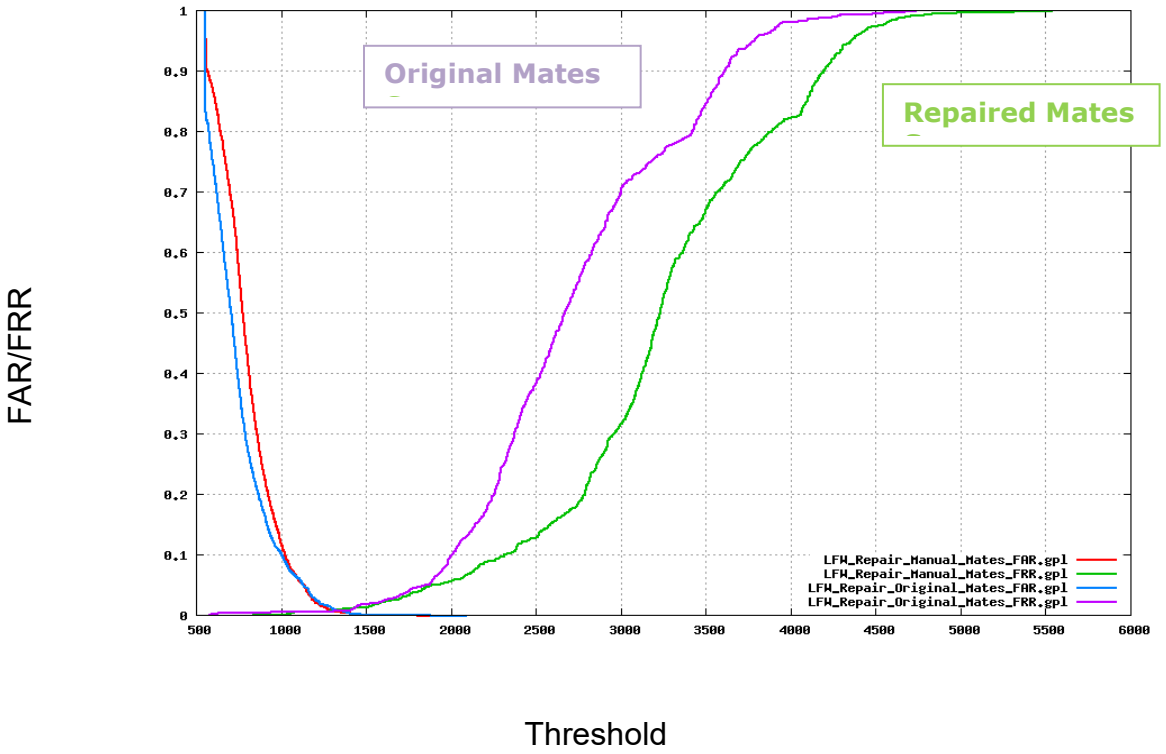

Figure 6: LFW Original/Repaired Mate/Imposter Scores (linear Y axis)

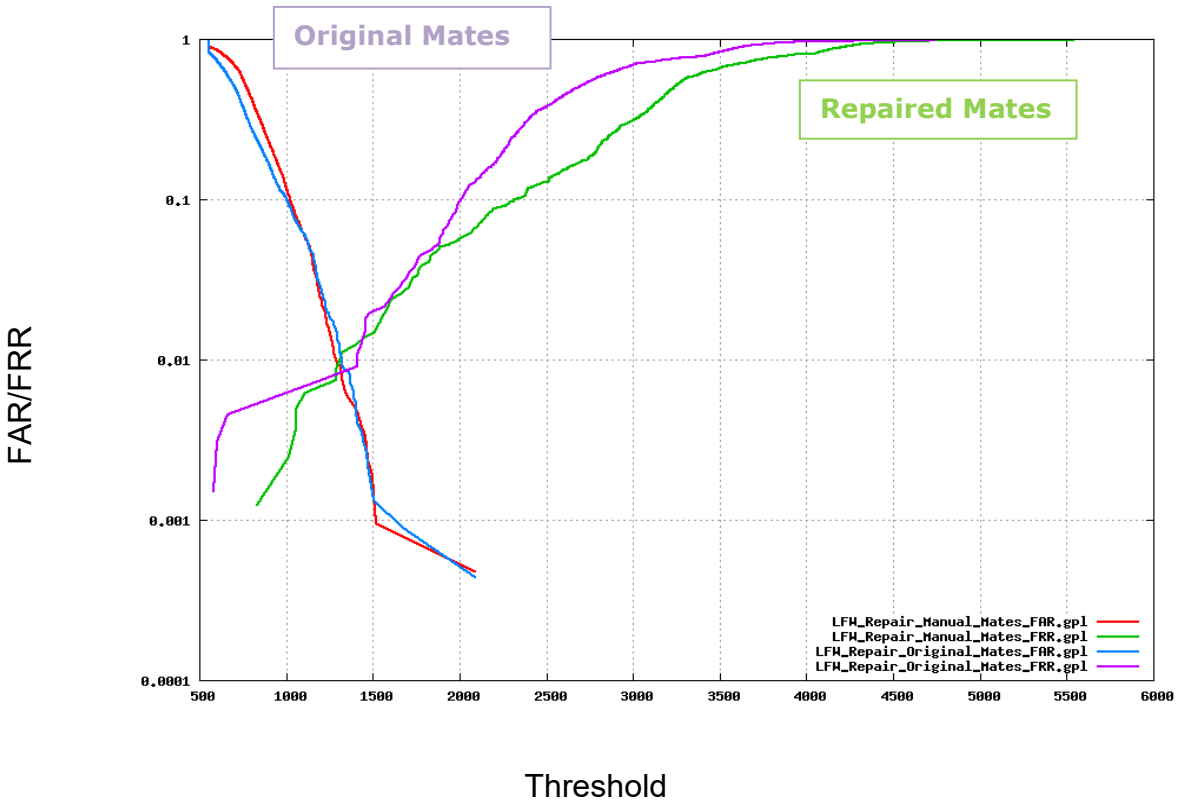

Figure 7: LFW Original/Repaired Mate/Imposter Scores (logarithmic Y axis)

Notes on Mate/Imposter scores:

- The mate scoring improved which was expected from the proper facial localization performed on the selected images.

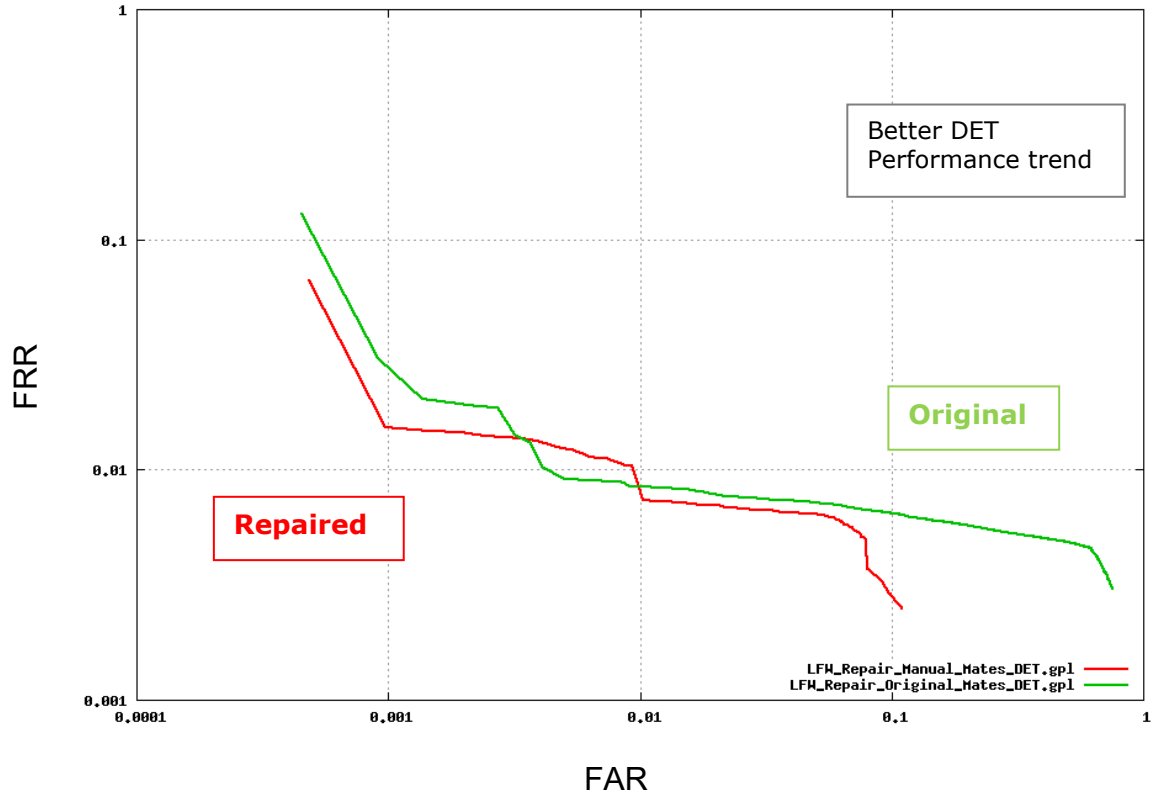

Figure 8: LFW Original/Repaired DET

Notes on DET:

- The repaired imagery showed improvement over most of the DET range plotted.

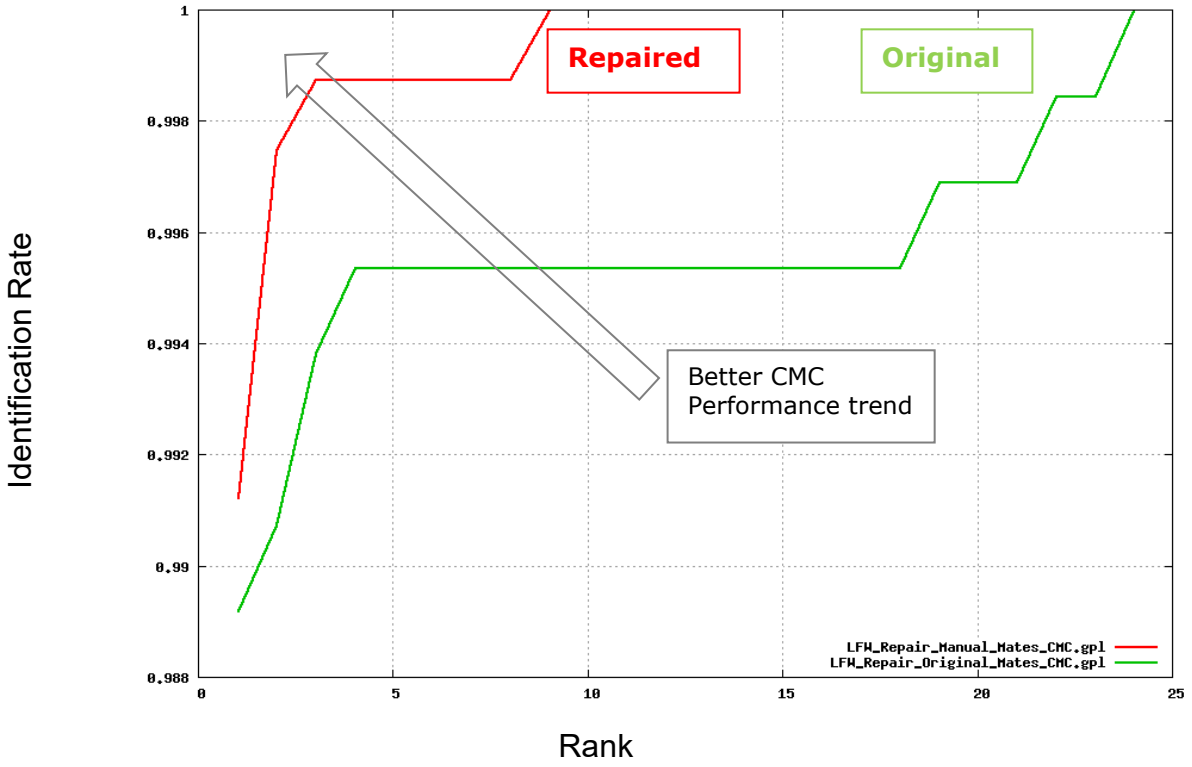

Figure 9: LFW Original/Repaired CMC

Notes on CMC:

- The repaired imagery showed improvement over the entire CMC plot.

Perhaps the largest outcome of this testing is presented in the CMC curves and can be described as follows:

- The original imagery has a 100% CMC point at a search results candidate of ~24.
- The repaired imagery has a 100% CMC point at a search results candidate of ~9.

Results:

- LFW gallery size: 13,277
- LFW images with quality less than 100: 182 (0.25% of the gallery)
- LFW images which were selected for manual facial localization: 32 (17.5% of the low quality images)
- CMC rank gain on the 32 images: ~15

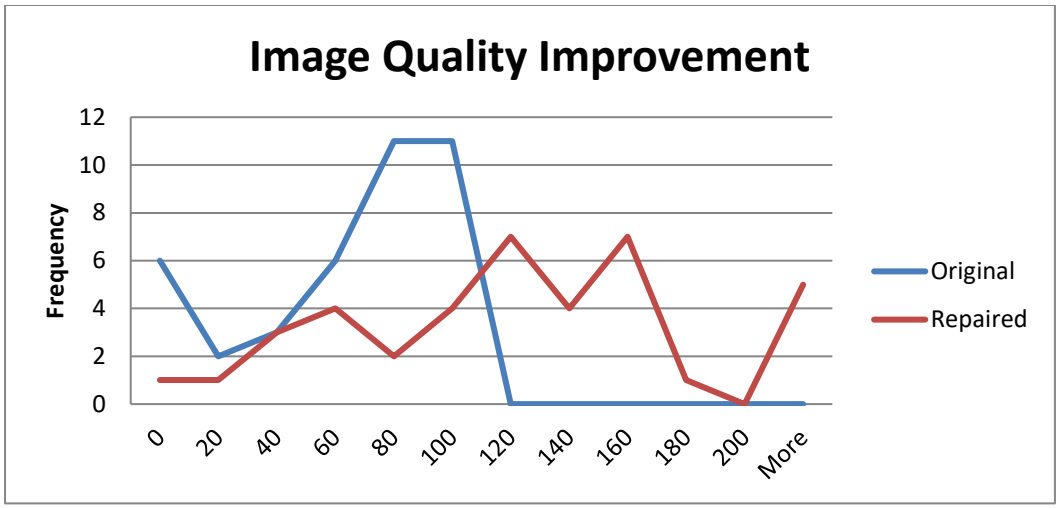

Figure 10: LFW Image Quality Improvement

Image quality improvement:

- Comparing the image quality of the original facial imagery vs. the repaired facial imagery shows an overall improvement in image quality

6.12 Outcomes

6.12.1 Based on this data set and the testing process documented here:

- Facial search performance is affected by facial image quality and if the facial algorithm properly localizes the face in the image.
    - For certain facial images, proper localization through manual alignment before templates are created can improve biometric usability.
- FAR, FRR, DET and CMC curves were utilized in these processes.
- Detecting and correcting low quality facial imagery which could be manually corrected can improve overall facial accuracy.
- Agencies need to determine their own assessment on whether their specific facial data and operational workflows can benefit from these processes.

FISWG documents can be found at: www.FISWG.org