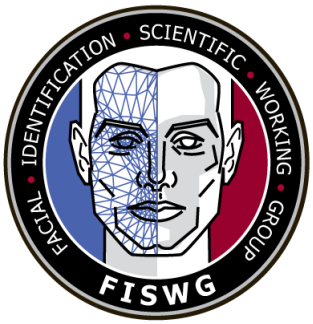

Disclaimer:

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: Lessons Learned in Investigative Searches

1. Scope

1.1 This is the final document in the FISWG Operational Assurance document series. The reader must have a working knowledge of this document series.

1.2 The scope of this document is to provide a list of lessons learned when doing investigative searches with facial imagery. This document will cover a wide range of issues all of which can improve the overall operational performance of an operational FRS deployment. This document is relevant to systems that operate with automated workflows as well as investigative systems requiring a human practitioner to review a candidate list.

1.3 An operational environment that performs investigative facial searching is not defined just by the assumed accuracy of the facial algorithm deployed in the biometric search core but needs to be defined by the overall support and attention to operational details of the complete operational procedures utilized by the agency deployment. Understanding how to ensure facial biometric accuracy is critical, but system management issues and support for the human practitioners is equally important.

1.4 Topics outside of this document include but are not necessarily limited to 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]

3. Terminology

3.1 *Definitions:* See ASTM E2916 Terminology for digital and multimedia evidence examination terms.

3.1.1 *Doppelganger:* an apparition or double of a living person.

3.2 *Acronyms*:

3.2.1 *CMC*: Cumulative Match Characteristic

3.2.2 *DET*: Detection Error Tradeoff

3.2.3 *FR*: Face Recognition

[1] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For Annual Book of ASTM Standards volume information, refer to the standard's Document Summary page on the ASTM website.
[2] ANSI/NIST-ITL Standard | NIST

3.2.4 *FRS*: Facial Recognition Systems

3.2.5 *FRVT*: Facial Recognition Vendor Tests

3.2.6 *GPU*: Graphics Processing Unit

3.2.7 *GUI*: Graphical User Interface

3.2.8 *IPD*: Interpupillary Distance

3.2.9 *NIST*: National Institute of Standards and Technology

3.2.10 *ROC*: Receiver Operating Characteristics

3.2.11 *ROI*: Return on investment

4. Summary of Guide

4.1 The document provides guidelines and techniques to help agencies that utilize automated FRS understand and consider holistic approaches that can be referenced to improve overall biometric performance and improved identity management.

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

4.3 The document is a continuation of the FISWG documents:

4.3.1 “Understanding and Testing for Face 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”

4.3.4 “Facial Recognition Systems Operation Assurance: Part 4, Manual Facial Localization”

4.3.5 “Facial Recognition Systems Operation Assurance: Part 5, Scoring Thresholds”

4.4 The issues presented in this document form a foundation for other considerations and applications when testing such as system setup and tuning. This document builds on these engineering principles but expands upon them to cover other operational areas that should be addressed to achieve a higher level of operational effectiveness.

5. Significance and Use

5.1 An operational deployment is not represented by a DET chart extracted from a NIST FRVT publication. An operational FRS deployment should be the result of a rigorous test and evaluation process as well as addressing a wider range of issues that all can affect the overall investigative performance of the agency Mission. Figure 1 illustrates the operational interdependencies as data is input into the system and practitioners finalize outcomes.

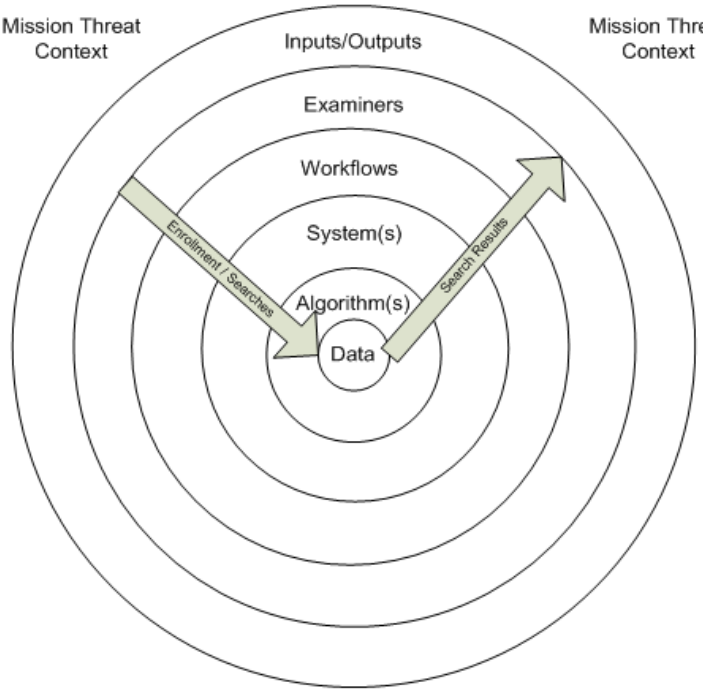

Figure 1 – FRS Operational Interdependencies

6. Topics

6.1 Proper Testing

6.1.1 As per NIST recommendations, testing for an FRS deployment should be conducted with agency data and algorithms. NIST recommendations should be expanded upon to include agency workflows, agency examinations procedures, human practitioner impacts, and the threat context within the agency mission.

6.1.2 Test data sets should have validated identity ground truth and properly replicate or represent operational gallery content and operational workflows.

6.1.3 Standardized test suites (i.e., test data set(s), select configuration(s), and test specific workflow(s)) should be created that can be easily repeated in operational and non-operational systems.

6.1.4 Test results and analysis should follow historical NIST testing:

- FAR/FRR: Critical for mate/imposter threshold analysis
- CMC: Critical for when evaluating identification rates and number of candidates for human review analysis
- DET: Overall performance for false accept/false reject
- ROC: Overall performance for true accept/false accept

6.1.5 Test results and analysis should include demographic differential performance, if applicable to agency-specific mission criteria.

6.1.6 The two most critical workflow parameters that result from operational testing are human review threshold score ranges and number of candidates. These should be defined by the system owner and implemented by the system administrator. If these aren't configured properly the results will include:

- Missed identifications
- Excessive human labor for little if any identification ROI

6.1.7 Understand algorithm limitations including:

- File sizes: minimum and maximum
- IPD: minimum and maximum

- Pose effects: yaw, tilt, roll
- Image rotation tolerances
- Effects of:
    - Excessive compression
    - Facial obstructions
    - Aspect ratio error
    - Motion blur
- Lightness and darkness limits
- Score normalization issues:
    - Gallery size
    - Gallery dependent or gallery independent
    - 1:N versus 1:1 scoring differences

6.2 Data Management and Workflows

6.2.1 Data Awareness and Gallery Management

6.2.1.1 Unintended consequences in biometric accuracy assessments will occur if:

- Data is enrolled regardless of image quality and practitioner usability.
- It is assumed that post enrollment corrections and adjustments to operational parameters and workflows can compensate for poor gallery management.

6.2.1.2 Vendor provided image quality metrics should be used to locate and possibly isolate poor quality imagery for potential repair.

6.2.1.3 Proper facial localization should be verified on all poses. If the face isn't localized properly the image quality metrics are not valid.

6.2.1.4 Images that are unusable or incompatible with the facial algorithm, e.g., a left or right profile image and an algorithm that is pose sensitive, should not be enrolled.

6.2.1.5 All FR algorithms have a low IPD limit where accuracy degrades quickly. Know what this limit is from proper testing.

6.2.1.6 If available, extract soft biometrics (e.g., sex, race, age, facial hair, etc.) for evaluation and potential usage.

6.2.1.7 Gallery content should be monitored. Image cleansing or repair should be performed if needed.

6.2.1.8 Subsets of the gallery may perform differently and need special attention to achieve desired accuracy levels. For example, images captured on film and transferred to digital will perform differently than images captured natively in digital format.

6.2.1.9 Areas of high and low recidivism in candidate search results should be identified. Excessive or improper manual image processing should be avoided as current algorithms are highly tuned to compensate for wide ranges of image quality. Manual image processing techniques should align with agency policies and procedures in addition to proper forensic principles.

6.2.2 Operational Metrics

6.2.2.1 Data awareness, controlled gallery content, and test results provide operational metrics that should be used to understand and adjust workflows with confidence.

6.2.2.2 The overall configuration and workflows may require refinements as the facial gallery expands.

6.2.2.3 Regardless of the galleries used for pre-deployment testing, once the system is deployed, the leading edge of the imposter curve (i.e., FAR) will increase in score. This may require adjustments to the human practitioner scoring thresholds.

6.2.3 Workflow Settings

6.2.3.1 Default search parameter settings should be derived from test results. The following key search settings should be available for case-by-case manual overrides:

- Image quality thresholds, if used. Thresholds for enroll and search may be different
- Number of candidates returned from searches
- Search filters, if used
- Algorithm configuration parameters, if available

6.2.4 Exceptional Cases

6.2.4.1 Outliers, such as low score mates and high score imposters, should be used to identify and mitigate operational gaps. These should be investigated to determine if

they are singular issues or indicative of broader systemic issues. For example, a higher occurrence of twins in the operational gallery than was used for pre-deployment testing.

6.2.4.2 If contractual and classification issues allow consulting with the algorithm provider may be warranted to mitigate operational gaps and address systemic issues.

6.2.4.3 Low score mate root causes include:

- Improper facial localization
- Poor image quality
- Pose variations
- Obstructions in the facial area
- Other image quality issues

6.2.4.4 High score imposter root causes include:

- Improper facial localization
- Poor image quality
- Twins
- Siblings
- Doppelgangers
- Other image quality issues

6.2.4.5 Using these exceptional cases can isolate gaps that could be closed or to improve operational awareness to others like it.

6.2.4.6 Reporting and consulting with the algorithm provider should be considered if contractual and classification issues allow this.

6.2.5 Search Filters

6.2.5.1 Filtering on sex, age, race, threat, behavior, location, etc. can be helpful when doing investigative searching:

- The candidates returned from the search will be of the context selected by the practitioner

6.3 Practitioner Considerations

6.3.1 Human Practitioner Value

6.3.1.1 System owners should continuously poll their practitioners for feedback (e.g., improved efficiencies, process improvements, operational gaps, system defects, suggested system changes, areas of concern, identified risks).

6.3.1.2 System owners should integrate this feedback into the strategic planning process and system improvements and upgrades.

6.3.2 Human Practitioner Impacts

6.3.2.1 Operational lighting considerations that may impact human practitioners include, but may not be limited to:

- Ambient light temperature control: Warm light, cool light, white light
- Ambient lighting variations that change by hour of day or time of year
- IT equipment considerations that may impact human practitioners include:

- Multiple monitors should be provided
- Monitors should meet photo-editing requirements:
    - 99% Adobe sRGB color gamut
    - Embedded color calibration
    - Peak color accuracy
    - Optimized color calibration
    - Hood
- Monitor color calibration should be configured to a uniform standard after polling practitioners
- Monitor selection should be accompanied with appropriate video card selection (e.g., embedded GPUs) that will sustain practitioner efficiency and throughput, and enable future capability expansion.
- System owners should be aware that monitors with too high a refresh rate, too large screen size, etc. may negatively impact the practitioners.

6.3.2.2 GUI design to enhance image analysis:

- GUI design should adhere to a “dark theme”, if possible
- Monitor background should be black

6.3.2.3 Eye strain:

- Frequent breaks should be taken when working on the computer to give practitioners time to “re-calibrate” their brain/eyes[3].
- Consideration of the 20-20-20 rule should be taken: every 20 minutes, look 20 ft away for 20 seconds.
- Consider computer eyewear designed for the precise distance between the eyes and the monitor, have anti-reflective coating, and a light tint

6.3.2.4 Practitioners should be assessed for color blindness, which is present in 7-8% of the human population. Agencies should evaluate if, and to what extent, color blindness affects their mission.

6.4 Facial Algorithms

6.4.1 Algorithm Provider

6.4.1.1 Operational support should be encouraged to ensure the product is functioning as advertised.

6.4.1.2 The relationships between FRVT tests and agency specific tests should be explained.

6.4.1.3 Contractual proposals should include technical support.

6.4.2 Algorithm Updates

[3] How Our Eyes See (kenrockwell.com)

6.4.2.1 NIST FRVT testing should be followed and any upgrades should be planned on a set schedule.

6.4.2.2 Algorithm updates or changes may affect the agency mission and workflows and should be followed by proper testing and developmental updates.

6.4.3 Pose Invariant Algorithms

6.4.3.1 Deploying a pose invariant algorithm can have a dramatic impact on an FRS if the facial images to be enrolled have variant poses.

6.4.3.2 There are many possible unintended consequences in going from an algorithm that does not properly process high yaw imagery to an algorithm that does:

- Facial localization may change
- Image quality aspects may change
- A larger gallery enrollment may occur
- More computational resources may be needed
- Changes to candidate lists may occur

6.5 Processes

6.5.1 Standards Groups

6.5.1.1 Participation in standards groups and reviewing of the documents produced by them can be highly beneficial and should be encouraged. These groups include:

- Facial Identification Scientific Working Group (FISWG): www.fiswg.org

- Organization of Scientific Area Committees (OSAC) for Forensic Science: https://www.nist.gov/organization-scientific-area-committees-forensic-science
- International Associations of Identification (IAI): www.theiai.org
- European Network of Forensic Science Institutes (ENFSI): www.enfsi.eu
- Other collaboration groups

6.5.1.2 Human practitioner evaluations should be encouraged.

6.5.2 Developmental Considerations

6.5.2.1 Recent states in the US have passed legislation for any new FRS which include:

- Algorithms must have been tested and reported via NIST FRVT.
- Accuracy testing mandates on agency data may follow.

6.5.2.2 Care should be taken to focus on iterative improvements while measuring incremental ROI. Examples include:

- Increase identifications and reduce misses by updating system configuration(s) and/or changing workflows
- Improve efficiencies, e.g., workflow process improvement

6.6 Key Takeaways

6.6.1 Considerations should be taken to optimize every step of the end-to-end process.

6.6.2 An operational workflow for investigative facial searching should be considered:

6.6.2.1 System: A facial image is received and is to be searched

6.6.2.2 Human: The image may be processed by a human practitioner before the search

6.6.2.3 System: The image is searched and a candidate list returned

6.6.2.4 Human: The candidate list is reviewed

6.6.2.5 Human: Potential candidates may be located

6.6.2.6 Human: The search results are finalized and an investigative lead may result

6.6.3 If any of these sequential steps are minimized, the overall impact to the entire process will be affected.

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