OSAC Technical Guidance Document

Framework for Implementing Passive Live Facial Recognition

https://doi.org/10/29325/OSAC.FI.0008

Facial & Iris Identification Subcommittee

OSAC
Organization of Scientific Area Committees for Forensic Science

OSAC Technical Guidance Document 0008

Framework for Implementing Passive Live Facial Recognition

Prepared for
The Organization of Scientific Area Committees for Forensic Science (OSAC)

Prepared by
Facial & Iris Identification Subcommittee, Digital/Multimedia Scientific Area Committee (SAC)
Organization of Scientific Area Committees for Forensic Science (OSAC)

January 2024

Document Disclaimer:

This OSAC Technical Guidance Document was produced using a consensus process within the Organization of Scientific Area Committees (OSAC) for Forensic Science and is made available by the U.S. Government. All OSAC members had an opportunity to comment on the document and provide suggestions for revisions on this OSAC Technical Guidance Document. The views expressed in the document do not necessarily reflect the views or policies of the U.S. Government. The document is provided "as is" as a public service, and the U.S. Government is not responsible for its contents.

Any mention of commercial equipment, instruments, or materials in this Technical Guidance Document does not imply recommendation or endorsement by the U.S. Government; neither does it necessarily imply that the materials or equipment identified are the best available.

Copyright Status:

Abstract

Technical advancements in both camera and facial recognition technology are such that live (Real Time) facial recognition is now technically feasible and there is increasing scope for use in several scenarios. The deployment of Live Facial Recognition (LFR) in a passive environment can provide rapid response to detect individuals predesignated as of interest to the deploying agency. The concept of and conditions of operation of an LFR system must be combined with Human-in-the-Loop decision making and underpinned by appropriate testing with regards system effectiveness. LFR systems must be deployed in a manner that balances the operational imperative with maintaining the anonymity of individuals with 'privacy by design' features enabled that give due regard to the right to privacy.

The understanding of these concepts will position agencies and governments to shape policy and governance to ensure responsible and ethical usage of this technology.

Keywords

Live facial recognition; Passive facial recognition; Design considerations; System Effectiveness; performance metrics; Ethical implementation

Table of Contents

List of Tables

List of Figures

Executive Summary

This technical document provides a framework for the implementation of a Passive live facial recognition system within the parameters of the described Concept of Operations. It provides an overview of live facial recognition and guidance on the design guidelines for implementation. **Central to the ethical implementation of a live facial recognition capability is the consideration of proportionality, human rights and the right to privacy**. This document describes 'privacy-by-design' features that should be implemented in support of maintaining people's anonymity. The document also provides information on the key performance metrics that describe the accuracy of a live (or Real Time) facial recognition system and guidance on how these should be measured. Finally, a number of recommendations are provided for those considering implementing such a system.

1. Introduction

Technical advancements in both camera and facial recognition technology are such that Live (Real Time) facial recognition is now technically feasible and there is increasing scope for use in a number of scenarios. LFR has a number of potential uses such as:

a) Supporting identification and arrest of people wanted for crime (fugitives, outstanding warrants, etc.);
b) Preventing people who may cause harm from entering an area; (this could include casinos with known parties working together or sex offenders from entering schools)
c) Supporting the identification of people about whom there is intelligence to suggest they may pose a risk of harm to themselves or others (e.g. stalkers, terrorists, missing persons, etc.)

This document is intended to provide information & promote understanding with respect to how these systems work, their implementation and optimization and how system effectiveness should be measured. **Critically, LFR systems can and should be designed & deployed in a way that adheres to the principles of proportionality, human rights, data privacy and ethical frameworks.** It is outside the scope of this Technical Guidance to document, in detail, these considerations but attention is drawn to a number of references that will provide appropriate guidance to organizations [1] [2][3][4].

This document aims to provide an overview of LFR, factors to consider when implementing a system and guidance on how to test and measure the effectiveness of a deployed system.

1.1. Concept of Operations (ConOps)

For the purposes of this document, LFR is defined as passive, automated, real-time searching of facial images from a video stream against a collection of reference images, referred to as a 'watchlist' in order to elicit an immediate response. The human operator is key for human-in-the-loop decision making to assess the alert and determine the appropriate response. There is no human input with respect to the submission of images from the video camera(s) however, human input is generally required with respect to building the watchlist.

This document focuses exclusively on live facial recognition as defined. Real-time controlled capture of co-operative subjects, for example at a gated access point such as E-gates is out of scope. The other main form of automated facial recognition referred to as 'Post-Event', 'Retrospective' or 'Forensic' (non-real-time searching of images against a database) is also out of scope.

The parameters of deployment for LFR will depend on the operational imperative and Federal, State or Local agency guidelines on the use of LFR [5][6][7]. Best practice dictates that where feasible, as many 'privacy by design' features are built in/switched on for each deployment.

1.2. Overview of LFR

1.2.1. High level data flow for LFR

The following workflow describes the ConOps above and incorporates a number of privacy enhancing features. Prior to any deployment, appropriate policy documents with checks & balances combined with an authority to operate should be in place.

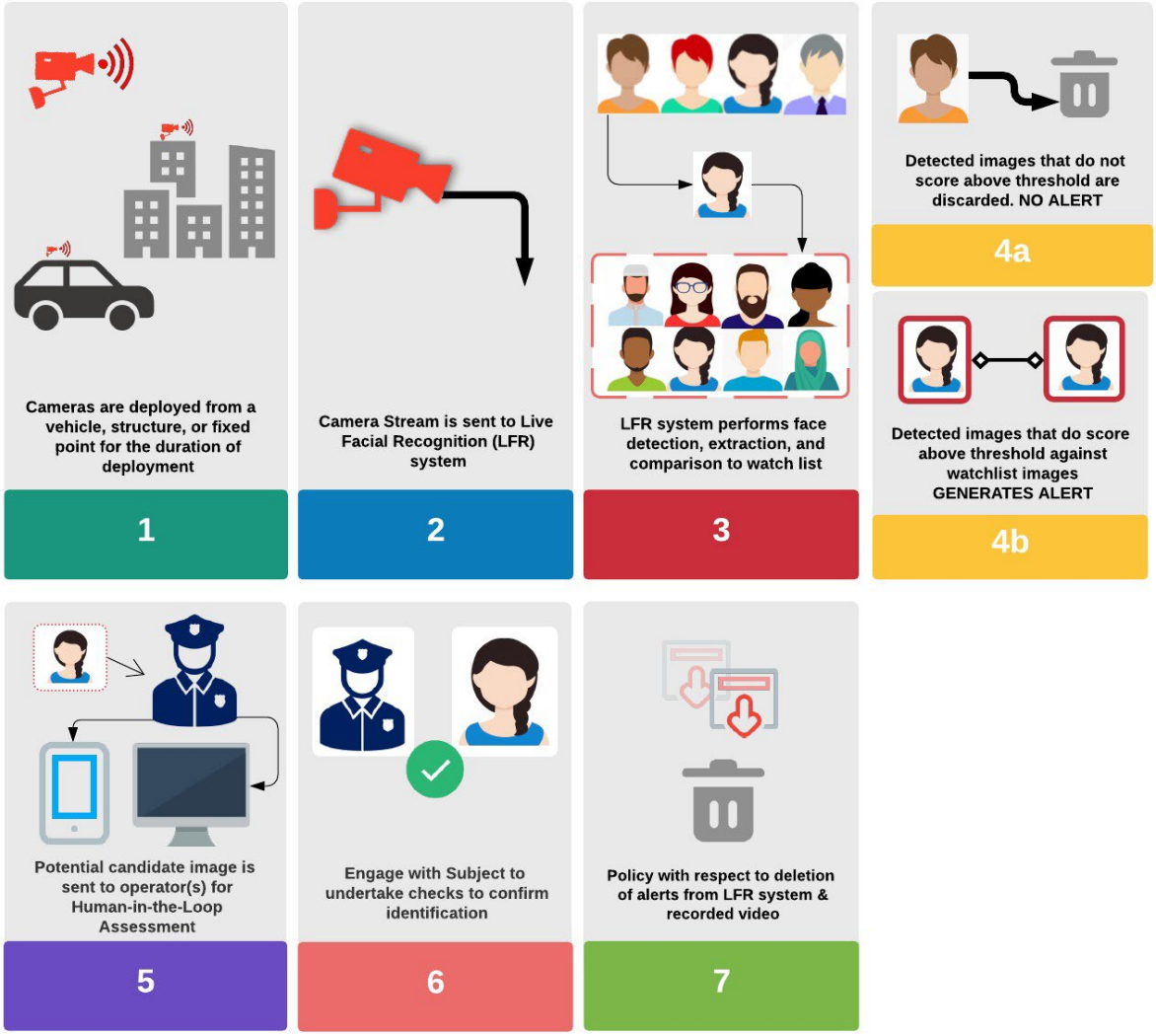

Figure 1 Concept of Operations for Live facial recognition

A set of reference images and associated metadata (for example missing persons or fugitives) is set as the 'watchlist', which is relevant to the operational imperative and has appropriate review, retention & deletion policies applied. As subject(s) pass an LFR camera, their faces are detected and converted into a mathematical representation, often referred to as a 'template'. The template is searched against the 'watchlist'. If a similarity score above

threshold is made between a subject and a watchlist image (referred to as potential candidate), the system generates a record or an alert of the potential match. **If no match is generated, all data relating to the subject including the image & the template is immediately & automatically deleted and no data of the subject is retained within the facial recognition system** [Privacy enhancing feature]. There may be legal requirements to retain the video associated with the LFR deployment and best practice dictates that a retention and deletion policy should be in place for retained video. Both the detected face from the video and the potential candidate from the watchlist are presented to the operator for human review and decision making. It is good practice to also include the full video frame image for review. **The faces of individuals within the context image that are not the subject of the alert should be redacted** [Privacy enhancing feature]. Potential steps in the decision-making chain, may include engagement with the subject who generated the alert & utilizing other methods to confirm identification.

1.2.2. Acknowledging concerns relating to use of Live facial recognition

A number of concerns have been raised regarding the use of LFR in public spaces by government or law enforcement. In order to build trust in the use of LFR, it is important that these concerns are acknowledged and addressed. It is equally important to highlight inaccurate statements or assumptions and note what an appropriately governed LFR cannot do. It is outside the scope of this document to fully address these concerns but they are described under three main categories for consideration.

- Myth 1 '*Live facial recognition is illegal*'

It may not be strictly necessary to develop a specific legal framework for the use of LFR or indeed any other technology. Instead, existing legislation can collectively provide a multi-layered legal structure to use and regulate the use of LFR [8].

- Myth 2 '*facial recognition is inaccurate and biased*'

It is incumbent upon the organization deploying facial recognition technology to undertake due diligence with respect to the facial recognition algorithm deployed. Not all facial recognition algorithms are equal or behave in the same way and testing undertaken by the National Institute of Standards & Technology has shown that generalized statements with respect to accuracy and demographic differential performance are not supported [9][10]. Testing undertaken under operational conditions for different use cases [11] has also shown that, at least for the specific algorithm tested, that:

  - There are settings the algorithm can be operated at where there is no statistical significance between demographic performance;
  - There was no demographic performance variation for Retrospective Facial Recognition; and
  - There was no demographic performance variation for Officer Initiated Facial Recognition

- Myth 3 '*LFR is intrusive and impacts on citizen privacy*'

In a properly implemented system, privacy is considered at every stage with appropriate governance and strong privacy-by-design built in. Through these measures, it is not possible to identify people who walk past the system if they are not on a watchlist. Images of subjects who pass the system are not collected for additional analysis and are not added to a

watchlist. Nor is it possible to track people as they go about their daily lives. The footprint of the technology should be relative to a specific operational use and location. Whilst passive LFR requires automated processing, decisions with regards to identity confirmation are made by the human reviewer and not by the LFR system. Only a very small percentage of people who walk towards an LFR system camera will generate an alert and not all of those alerts will result in an engagement with a law enforcement officer or other official [Figure 2].

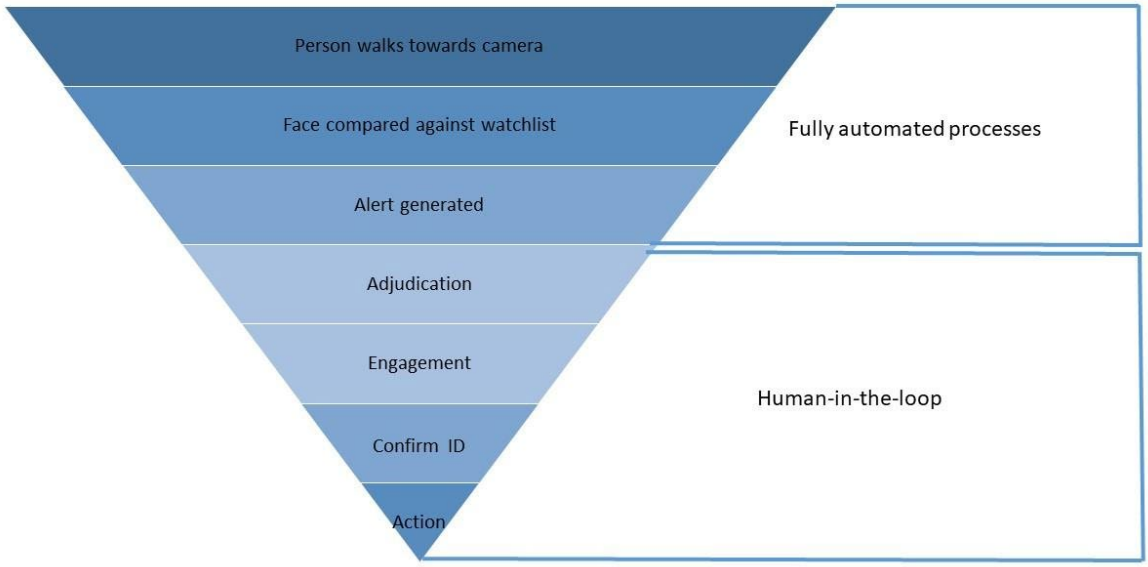

Figure 2 The process in filtering recognition opportunities to identify a person on the watchlist

2. Design Guidelines

Effective live facial recognition of 'passively-imaged' subjects in a semi or uncontrolled environment requires dedicated design effort [12]. Watchlists need to be of sufficient quality images with quality standards applied at the time of enrollment. A human-in-the-loop assessment stage is critical to adjudicate alerts generated by the system. From an operational perspective, there should be sufficient resources to respond to and deal with alerts against a watchlist. If the objectives are too broad and the system not correctly optimized or implemented, the amount of human resources required to respond to alerts may be prohibitively high.

For a recognition service (which includes the hardware, software, the system operator and associated resources on the ground) to deliver the desired results, all components need to be optimized and inter-operate correctly.

A live facial recognition system will consist of many components. For example, many facial recognition systems are plug-ins to Video Management Systems (VMS) that communicate

directly with cameras, video archival etc. Those components that do not directly relate to the successful use of facial recognition are not considered in this document, although due attention should be paid to governance and procedures where, for example VMS infrastructure may be needed to ensure video is properly archived for auditing / legal compliance. Directly relevant components include:

- Cameras, their placement in the deployment area, resolution, lens, dynamic range, On-edge Processing and network interface
- Network architecture to ensure sufficient bandwidth for data transmission
- facial recognition software that detects faces in the video stream, converts the facial images into templates, compares these against the watchlist and provides information on the results of the search in the form of an alert to an operator
- Database of reference images and associated metadata, collectively referred to as the watchlist
- An operator who assesses the alert and determines the appropriate course of action;

2.1. Cameras and their placement

Cameras should be selected so that the image resolution, frame rate, field of view and low-level light performance can provide images of sufficient quality for use in the automatic facial recognition application. Inter Eye Distance (IED) is a critical factor in successful operation of an LFR system. Current FR systems typically work better with a facial image that has between 64 to 128 pixels between the center of the subject's eyes. Under ideal environmental and operating conditions IED could be at the lower bound but at the same time, non-ideal conditions may require a higher IED. Optimal IED is very much algorithm dependent and the FR vendor should advise on specific requirements for their system.

NIST reports [9] provide a good source of information on relative dependence for IED, but ultimately these are not determined under the same set of conditions that an LFR system may be deployed. As such, it is imperative that IED considerations are tested under expected operational conditions.

Ideally the environment should be managed such that every face is evenly illuminated. Highly directional lighting, for example strong sunlight, should be avoided, which may require consideration of how the lighting will change throughout the day. The cameras should be operating with a Wide Dynamic Range in order to generate sufficient quality images under a variety of lighting conditions.

Cameras should be positioned to capture faces as close as possible to a frontal pose although the tolerance for detection and recognition of off angle faces has increased and the latest high performing algorithms can successfully identify subjects at increasing amounts of yaw and pitch pose variation. Frontal pose capture typically requires the cameras to be much lower than is normally the case for legacy camera systems. In

general, cameras mounted at between 1.83m and 2.44m provide optimal capture conditions [13]. Tolerance to off angle facial images is very much algorithm dependent and therefore LFR systems must be tested under operationally realistic conditions.

The zone of recognition is defined as that zone within the total field of view of the camera within which the conditions for facial recognition are optimized. In general, the zone will be smaller than the field of view of the camera; for example, not all faces in the field of view may be in focus and not every face in the field of view will be imaged with the minimum necessary Inter-Eye Distance (IED) [Figure 3].

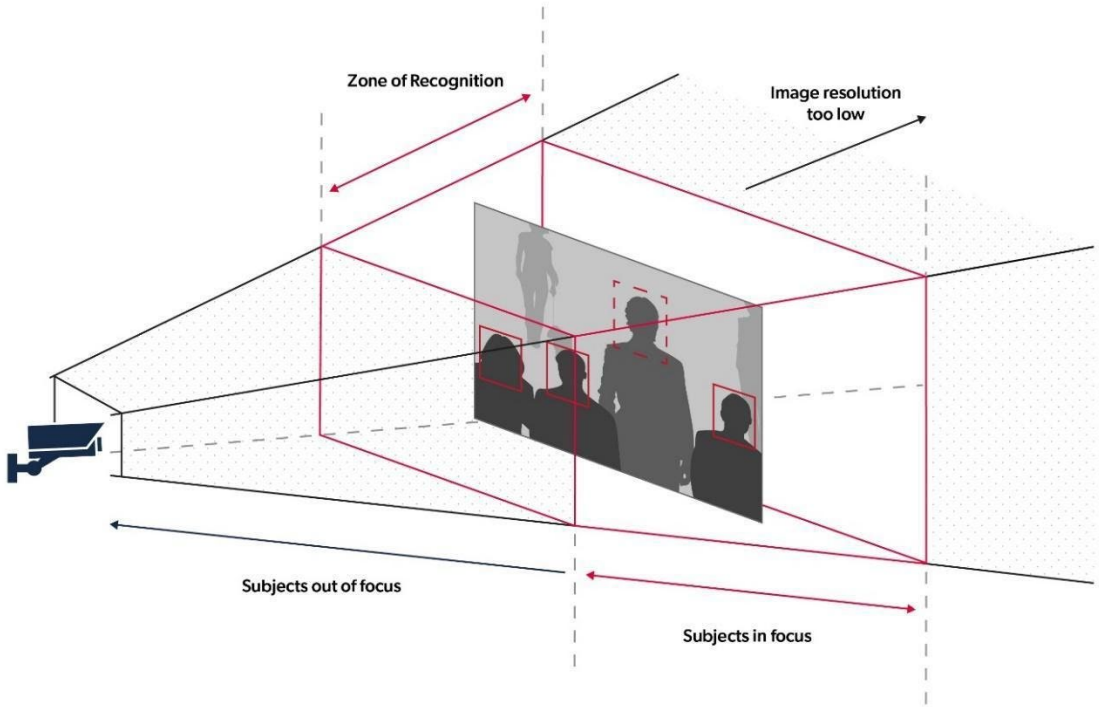

Figure 3 Zone of recognition

The camera resolution defines the limit of the horizontal capture space that meets the requirements for face recognition. With a fixed IED pixel requirement and a known camera native horizontal resolution, the scene capture width can be calculated.

Typically, adults have an Inter Eye Distance of approximately 64mm (see Figure 4).

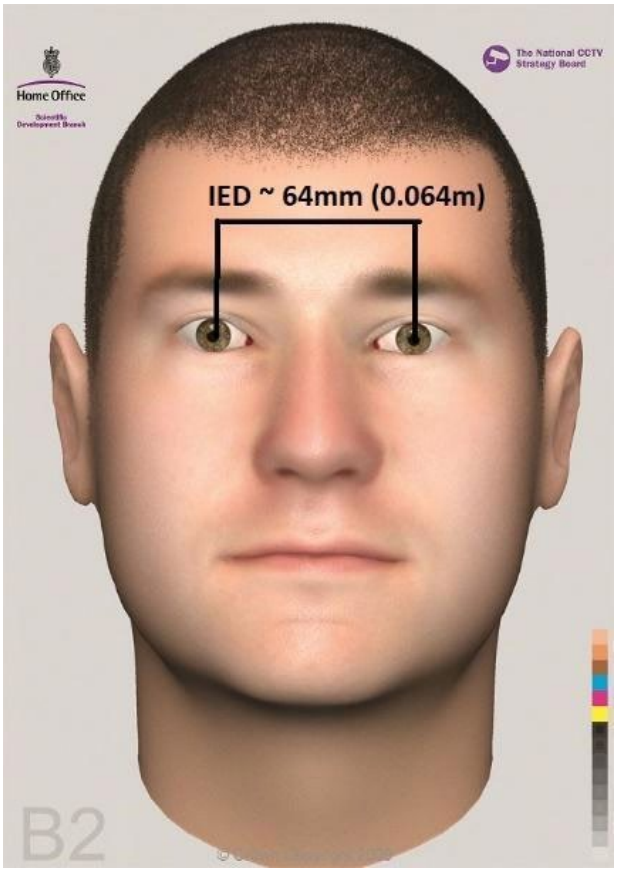

Figure 4 Typical Inter Eye Distance (IED) is circa 0.064m) [14].

The pixel density minimum requirement for facial recognition under non-ideal conditions varies by algorithm but is typically between 64 - 128 pixels IED. Taking 80 pixel IED as an example;
This equates to a pixel density of 1.26 pixels per mm (80/64), or 1260 pixels per m.
To achieve this level of pixel density a 2Mpixel camera (1920 x 1080 pixel frame size) will give a horizontal coverage of 1.53m (1920/1260).
Similarly, the horizontal scene capture width can be calculated for other IED pixel requirements and camera resolutions as set out in [Table 1].

Table 1 Horizontal scene capture width for facial recognition

Camera Resolution	IED (pixels)	Horizontal Capture width (m)
2MP (1920 X 1080)	64	1.92
	80	1.53
	128	0.96
5MP (2560 x 1920)	64	2.5
	80	2.03
	128	1.28

On the basis that no compression is applied to the camera stream and all the available resolution is streamed to the LFR system, a typical 2 MP camera will provide sufficient resolution for LFR to work on between three and five people standing side by side. Therefore, consideration needs to be given to camera location and the physical environment to manage the flow of people within the zone of recognition and to direct their gaze to a frontal pose. Caution needs to be exercised because if the flow is reduced beyond a certain level, individuals may be grouped very close together, occluding or partly occluding the faces of the people behind them.

The use of an attractor to direct the subject's gaze towards the camera may help to obtain better quality images and thus improve recognition rates. In 2017, NIST published the 'Face In Video Evaluation (FiVE) Facial Recognition of non-cooperative subjects [13] and showed that the False Negative Identification Rate (or the miss rate) was reduced through the use of such attractors. Examples of an attractor include a 'digital mirror', which is a monitor that displays the camera view and is positioned so that people can see themselves walking towards the camera or an agent audio video, which is a monitor displaying a moving avatar with associated audio. Across all algorithms tested the reduction was greatest using the agent audio video but the magnitude of the reduction was algorithm dependent, with the highest performing algorithms (with relatively low miss rates to start with) having the least relative impact.

2.2. Network Architecture

The network architecture will depend on the Concept of Operations including the number of concurrent live camera feeds, the control mechanism of the watchlist and the immediacy requirements of the response time to an alert. Options are detailed in [Table 2]

Table 2. Methods for measuring the number of recognition opportunities

Architecture	Pros	Cons
Closed system with cameras connected either directly hardwire or by encrypted wireless to co-located FR server	Supports Rapid Deployment to site specific location Supports immediate response in the field	Processing server hardware required for each location
Remote cameras and centralized FR server	Ease of centralized management of watchlist Ease of adding more cameras to the system Ease of scalability	Processing load for full frame rate video* Speed of response back to field

*Processing speed and load can be mitigated through the use of On-edge processing. There are a number of options and factors to consider

- 'Smart' cameras with inbuilt Face Detection and template generation capability
- Hardware or software based camera agnostic face detection that is co-located with the camera. This takes the full video stream from the camera, detects the face and only the detected face is transmitted to the FR server.

Both options reduce the bandwidth required to transmit and the processing load on the back end server. However, caution should be exercised and consideration should be given to:

- The compatibility of the template generated by a smart camera with a back end facial recognition system
- The effectiveness of the Face Detection algorithm – is it as good as the native Face Detection built in to the facial recognition system
- Quality – the parameters set for Face Detection may not be sufficient for facial recognition. This will result in faces being sent to the FR server that cannot be 'enrolled' for searching. This may have an impact on 'privacy by design' features built in to the system as it may result in a cache of detected faces that are not searched against the watchlist.

2.3. Facial recognition Software Configuration

Most FR systems allow the user to adjust a number of parameters including:

- the maximum number of faces to detect per frame and the framerate

- the decision threshold score for recognition; and
- how the system deals with multiple 'alerts' against the same individual

2.4. Relationship between processing throughput & system configuration

For a given camera feed, a number of factors influence the amount of processing required, including:

- The number of cameras feeds being processed
- Resolution of the camera
- The number of Frames Per Second (FPS) processed by the system
- The minimum resolution bounding box for detection and
- The number of faces in the field of view.

Detection and processing of faces is a computationally intensive task. High resolution cameras may generate a processing load that is difficult to sustain in the field if you are trying to detect low resolution faces. If the system is set to process too many faces, this may result in a delay in the system response. It may also result in missed alerts due to 'dropped frames' where the software skips some of the video footage in an attempt to 'catch up'. Ultimately, more processing power can address these issues but these considerations need to be determined prior to the operational deployment of the system and there may be operational constraints on this.

The LFR system itself will have two primary computational bottlenecks when processing streaming video: face detection (i.e., finding faces present in the video) and face representation (i.e., generating feature vector templates that can in turn be used for face recognition).

The FPS of the system directly impacts both face detection and face representation throughput. While most cameras output 25/30 FPS, most LFR systems only need to process between 5 and 20 FPS. The number of FPS will depend on the environment with typically;

- a controlled slow moving turnstile/gate people flow requiring 5-8 FPS
- a medium flow queue requiring 10FPS and
- a fast-moving uncontrolled people flow (for example at a transport hub) requiring 15 – 20 FPS.

The vendor should be consulted when choosing the FPS processed by the LFR system, but it should be noted that there is a direct relationship between the number of cameras being processed, the FPS of those cameras and the number of CPU or GPU cores needed to process streaming video. For example, a 4 FPS system will generally require 2x more computing resources than a 2 FPS system.

Factors that influence face detection throughput (aside from FPS and number of cameras) are the resolution of the camera and the minimum resolution bounding box for detection. Unless the system is designed to only detect a pre-specified number of faces, then, in general, face detection will not be impacted by the number of faces in the field of view.

However, the number of faces in the field of view does impact (along with FPS & number of cameras) the face representation speed. Each detected face in the field of view will in turn result in a template being generated. There is (generally) a linear relationship between the number of templates being created and the amount of CPU or GPU cores needed to perform this face representation step. If a camera is only expected to see a maximum of 2 faces at a time, as opposed to 20 faces at a time, then significantly less hardware will be required for generating searchable face recognition templates. The vendor should ultimately be consulted for information on computing hardware needed for the expected number of faces being processed at a time. The NIST Facial Recognition Vendor Test (FRVT) [9] reports "Template Generation Speed" which is a highly informative metric for this consideration.

2.5. Decision threshold score for recognition

There is a trade-off between the true recognition rate and false alert rate as shown in [Figure 4]. Live facial recognition requires that a 'similarity' threshold is set before a potential match alert is generated. Setting this decision threshold is a critical step as adjusting the decision threshold downwards may increase the True Positive Identification Rate but can have the effect of increasing the number of (false) alerts that must be dealt with. Likewise, adjusting the threshold upwards can mean that subjects who are on the watchlist might walk past the camera but not generate an alert.

A good starting point is the default threshold recommended by the system vendor. However, it is recommended that additional scenario or operational evaluation is undertaken to ensure that this threshold meets the Concept of Operations of the deployed system (see section on Key Metrics for measuring LFR effectiveness).

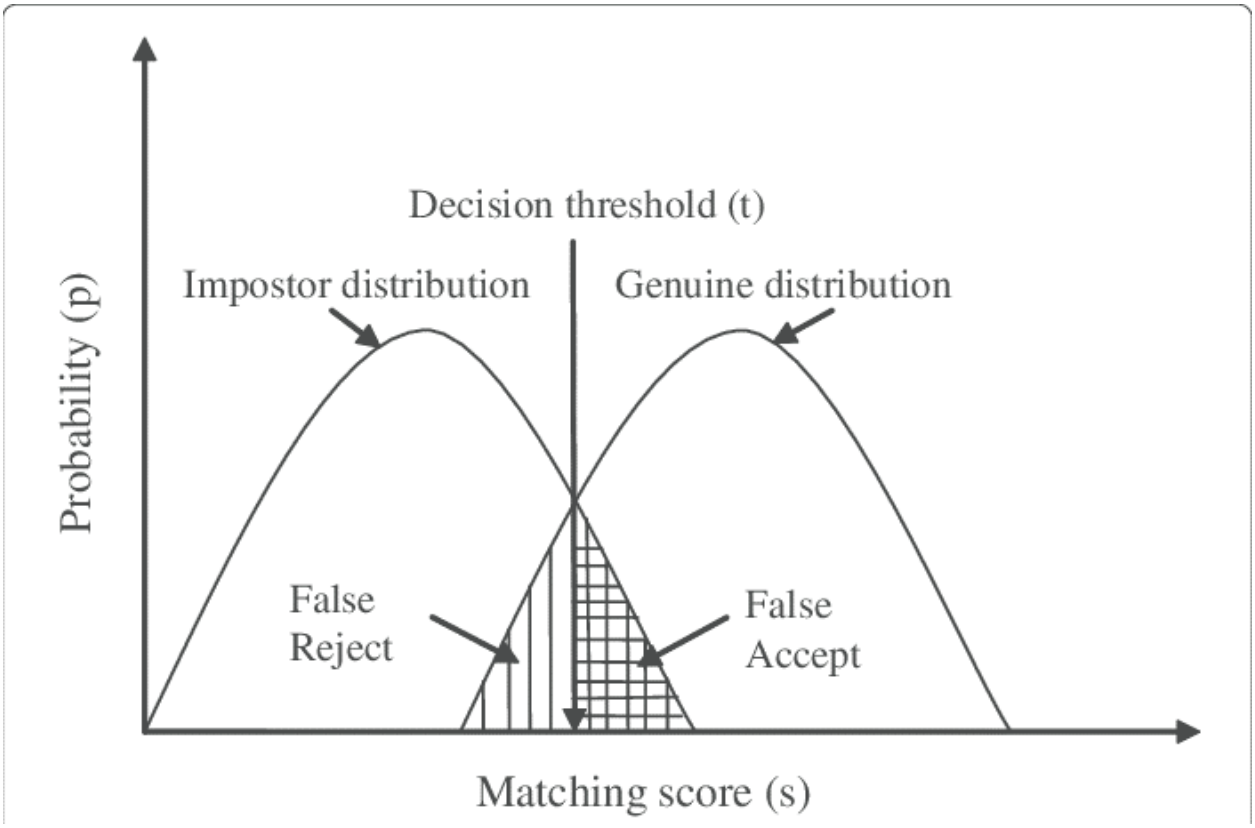

Figure 5 Impact of threshold ('matching score') on the number of False Rejects & False Accepts

2.6. Multiple 'alerts' against the same individual

In order not to overload operators with multiple alerts within a short space of time, when there is an alert (match score above threshold) between an individual and a watchlist subject, additional or repeated alerts for that individual should be suppressed for a configurable period of time to allow the individual to clear the Zone of recognition. The LFR system should automatically track each person within the operational zone and thus suppress redundant alerts, but the presence of this functionality needs to be confirmed with the vendor.

There are a number of options to consider here, not all of which are available in every LFR system:

- The first image of the individual that generates a score above threshold generates the alert
- The first image of the individual that generates a score above threshold is 'tracked' and if a subsequent image generates a higher score, until the individual exits the Zone of recognition, that image generates the alert
- A short video is created by a configurable number of frames backwards & forwards from the first image of the individual that generates a score above threshold

Each option has pros and cons and the selection will be dependent on the operational requirements and any constraints of the deployment.

Under the first condition described, the comparison score logged for each alert is that of the first recognition opportunity that scores above threshold. This may not be the highest comparison score possible for the recognition opportunity and as a consequence, the effect of using a higher decision threshold cannot be inferred from the logged comparison scores. This might also not be the best quality image from a human review perspective.

Under the second condition described, this should provide a better image of the individual to the operator for review purposes, but consideration needs to be given to allowing sufficient time between alert & the subject existing the Zone of recognition to facilitate engagement.

Under the third condition described, this will provide more material to the operator to compare against the watchlist image, but may increase the time required to review the alert.

2.7. Watchlist

The 'watchlist' consists of facial images and associated metadata of subjects. The quality of the watchlist image is key to performance of the LFR system and it is recommended that, as far as possible, watchlist images meet the standard for mugshot or passport images as set out in [15]and [16]. The size of the watchlist will be dependent on the operational imperative and most LFR systems will handle watchlists in the order of thousands. Consideration should be given to including more than one image of a subject within the watchlist as studies [13][17] show a decrease in the False Negative Identification Rate through enrolling multiple images. One of these images will have to be selected as the 'master' image to be returned as part of an alert. Generally, it is best practice for this to be the most recent image of the subject.

Functional features of an LFR system should include the ability to create different watchlist partitions, configure the alert threshold and the alert response by subject or by watchlist. Notifications or alerts to operators should clearly distinguish between, for example, subjects who are on a missing or vulnerable watchlist and subjects who are on a crime watchlist.

2.8. Operator Assessment

Live facial recognition requires a 'human-in-the-loop' to assess the alerts generated by the system.

Alerts should consist of the localized facial image, the watchlist image and metadata associated with the watchlist subject. A context image that shows the full video frame is very useful and helps operators locate the person who generated the alert.

There are multiple papers on human ability for non- familiar face matching[18][19] and on training requirements for operators undertaking assessment [20]. It is outside the scope of this paper to reiterate the detail of these papers here, except to note that the human is a key element of the end-to-end process of LFR & consideration must be given to selecting suitable candidates and to their training, which is comparable to the facial assessor role as defined by FISWG [21].

3. Key Performance Metrics

The overall application accuracy of an LFR deployment accuracy can be considered to consist of the combined LFR technology accuracy and the human in the loop decision-making process. As such, both should be measured.

3.1. Description of technology metrics

Standards mandate reporting performance of identification systems in terms of the frequency of **two** error conditions of the identification process; false negative and false positive rates [22] and [23]. The error rates should be measured over recognition opportunities, i.e. the period that a subject is walking through the Zone of recognition. Therefore, it is incorrect to describe the 'accuracy' of a live facial recognition system by a single figure (e.g 98% (in)accurate). It is more appropriate to describe system effectiveness in terms of the two metrics described below.

3.1.1. False Negative Identification Rate (FNIR)

The False Negative Identification Rate (FNIR) is the proportion of recognition opportunities of subjects who are on the watchlist which don't generate a correct alert.

$$FNIR\ (N,T) = \frac{Number\ of\ recognition\ opportunities\ by\ subjects\ on\ the\ watchlist\ not\ generating\ a\ correct\ alert}{Number\ of\ recognition\ opportunities\ by\ subjects\ on\ the\ watchlist\ *}$$

where *N* represents the size of the watchlist, and *T* the threshold that the comparison score must exceed for an alert to be generated.

*Or in other words, the number of times a subject enrolled in the watchlist appear in the video sequence

FNIR states the "miss" rate. Sometimes it is preferred to talk in terms of "hit" rates. The complement of FNIR is the **True Positive Identification Rate** (TPIR).

TPIR(*N,T*) = 1– FNIR(*N,T*).

3.1.2. False Positive Identification Rate (FPIR)

The False Positive Identification Rate (FPIR) is the number of individuals who pass the LFR system but are not on the watchlist and who (incorrectly) generate an alert as a proportion of the number of people appearing in the video sequence

$$FPIR\ (N,T) = \frac{Number\ of\ false\ alerts\ for\ subjects\ not\ on\ the\ watchlist}{Number\ of\ recognition\ opportunities\ for\ subjects\ not\ on\ the\ watchlist}$$

where *N* represents the size of the watchlist, and *T* the threshold that the comparison score must exceed for an alert to be generated.

Note – The denominator in the FPIR includes only the instances where individuals not on the watchlist are scanned by the system.

There is significant confusion over the proper interpretation of these metrics and these interpretations tend to focus solely on the number of alerts generated. However, by doing so, they ignore the significant volume of correct decisions that the system makes when it does not generate an alert against a subject who is not on the watchlist.

Additionally, the prior probability of a watchlist subject being present may be relatively low and therefore, even for high performing systems, it may transpire that the number of false alerts might outnumber correct alerts. See [Figures 6, 7 & 8] for a pictorial representation of this.

Note: The percentage TPIR & FPIR rates given below are examples only and should not be taken as indicative of any particular system performance.

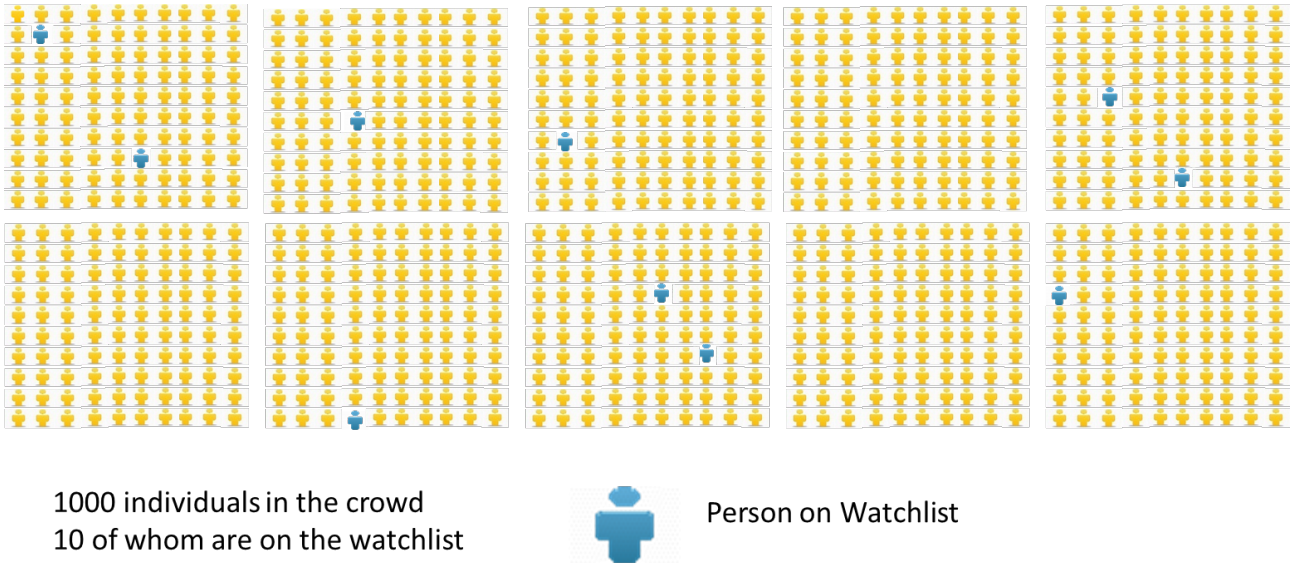

Figure 6 Every person who passes the LFR system generates a recognition opportunity

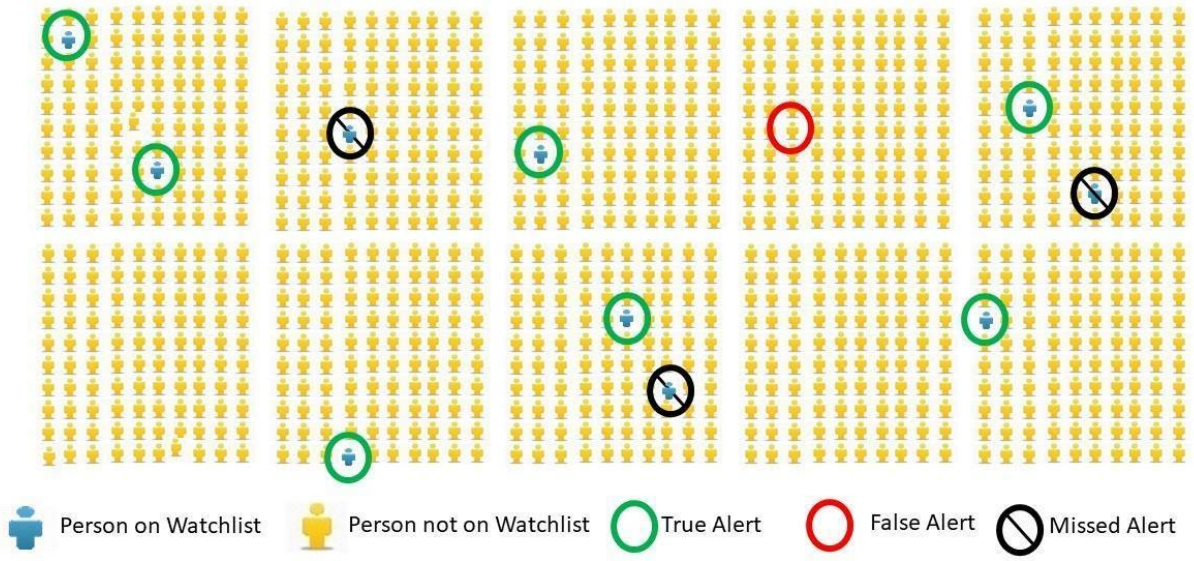

Figure 7 Visualization of hypothetical True Positive & False Positive Identification Rates

The **True Positive Identification Rate** would be 70% if 10 people on the Watchlist pass the LFR system, and a Correct Alert is generated for 7 out of 10 of those people (with no Alert being generated against 3 of those people – Missed Alert).

The **False Positive Identification Rate** would be 0.1%, if for every 1,000 people that passed the LFR system, an Alert was generated against one person who was not on the Watchlist (simplified to demonstrate the concept of TPIR & FPIR)

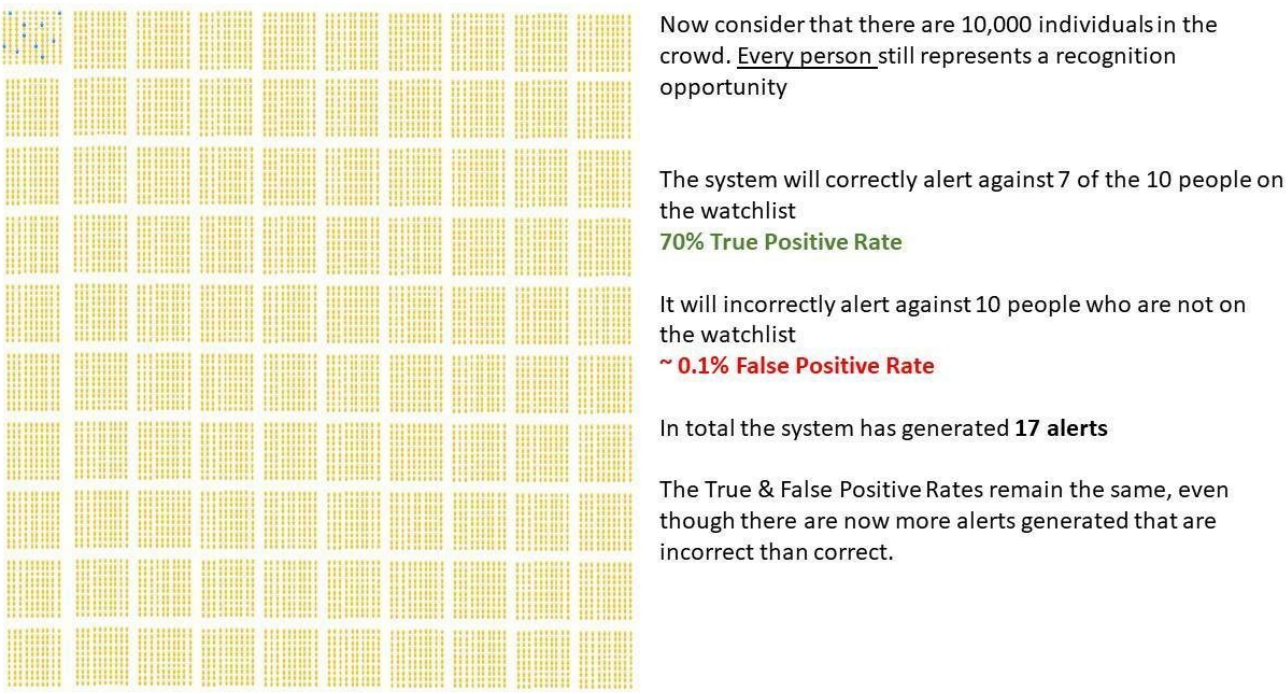

Figure 8 The impact of the number of transactions on the number of True and False positive alerts

A key concept of LFR metrics is that it is the rate of true and false positives that should be measured, not the number of true and false positive alerts.

System tuning is critical to balance the true positive and false positive rates. It is possible to tune the LFR system such that you decrease the False alert rate but this can/will have a corresponding effect on the True Positive Rate. The converse is also true. The operating point of the system will be dependent on the concept of operations and operational conditions.

3.2. Demographic Differential Performance

In 2019 NIST published the first report on large scale testing of demographic effects on FR accuracy[10]. In this report NIST defines differential performance as a "difference in the genuine or imposter [score] distributions" of similarity scores computed over a collection of images from two (or more) demographic groups". It is outside the scope of this paper to document how to determine demographic differential performance for a LFR deployment, except to note that three different cohorts must be taken into consideration, the demographic makeup of;

- Subjects on the watchlist
- Subjects who generate recognition opportunities

- Subjects on the watchlist who are physically present and generate a recognition opportunity.

The Human-in-the-Loop assessment of the alerts is essential and they, not the system, make the final identity and appropriate action decision.

3.3. Human in the Loop decision metrics

A suitably trained person [Reference Placeholder – awaiting FISWG training guidelines publication] must adjudicate every alert that is generated by the LFR technology and make a decision as to the appropriate action to take. There are a number of possible outcomes from each adjudication as set out in Figure 9.

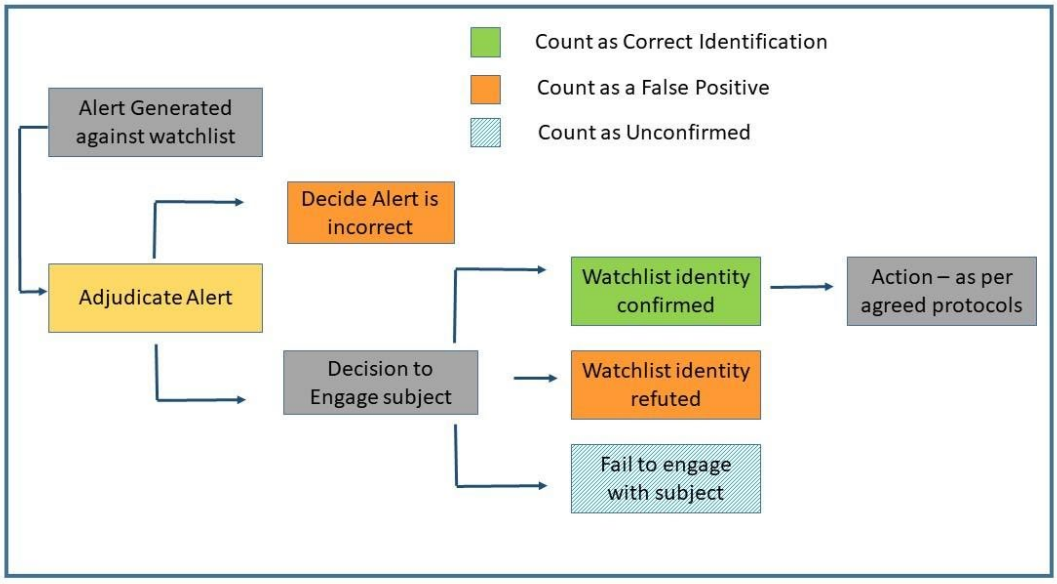

Figure 9 Possible outcomes for each system generated alert

It should be noted that the Ground Truth of alerts generated against an operational watchlist is not known. Therefore, the False Positive rate may be recorded as higher than it actually is as an operator may adjudicate that a system generated alert is incorrect, when in fact, it is correct. Additionally, there is a potential outcome where an operator adjudicates an alert to be correct, but fails to confirm the identity of the subject who

generated the alert. As it is impossible to determine if this is a False Alert or a True Identification, then the most appropriate way to classify this is as unconfirmed.

4. Determination of LFR technology metrics

Best practice dictates that ground truth data should be used to test biometric systems and determine the LFR metrics. However, this might not always be feasible in an experimental setting because of the high footfall required to generate sufficient transactions to measure the False Positive Identification Rate. The ground truth may not be fully known when using operational data and therefore, consideration must be given to the magnitude of uncertainty in a metric implied by incomplete ground truth [13]. Additionally, an assessment of the uncertainties in the estimates of the FPIR, TPIR & FNIR should be considered and are particularly important in trails where, for example, no false positives are observed. There are several approaches to this, so the exact method to select is outside the scope of this document.

Determination of the **True Positive Identification Rate** is made based on recognition opportunities by subjects known to be present, as there is no way to count the number of people on the operational watchlist that are missed by the LFR. A number of volunteer subjects should be seeded into a watchlist and seeded into the flow of people who pass the LFR system. This group of subjects is commonly referred to as a 'Bluelist'. These subjects may generate multiple recognition opportunities during the course of a deployment and;

$$TPIR = \frac{Number\ of\ correct\ 'Bluelist'\ alerts}{Number\ of\ Bluelist\ subjects\ recognition\ opportunities}$$

It should be noted that images of Bluelist subjects should be seeded into the full watchlist and that Bluelist subjects are compared against the totality of the watchlist, not just the Bluelist partition.

The False Positive Identification Rate requires knowledge of the total number of recognition opportunities.
Table [3] sets out methods to measure the number of recognition opportunities, all of which are estimates and have pros & cons.

Table 3 Methods for measuring the number of recognition opportunities

Method	Pros	Cons
Manual count of the number of subjects who pass within the Zone of recognition for entire duration of deployment	Accurate count	Resource intensive Subject to human fatigue & error

Utilize the 'Face detected' count of the LFR system	Automated – part of the system logs	The system will only count the subjects it 'sees', not providing an accurate count of recognition opportunities If the FR system tracklet is disrupted (by person momentarily moving out of view) then system will generate two counts for the same subject
Sample time periods during the deployment & undertake manual count of the number of subjects who pass within the Zone of recognition	Less resource intensive than manually counting for the full duration of the deployment Not subject to system errors	The sample duration & periodicity needs to be carefully considered to take natural peaks & troughs into account
Utilize one or more LFR system independent 'flow' count algorithms	Automated process Can average count if utilizing multiple systems	The system will only count the subjects it 'sees'

Once the number of recognition opportunities has been calculated, which can be undertaken retrospectively if a full frame rate recording is taken from the camera streams and with removal of data from bluelist recognition opportunities, the False Positive Identification Rate (FPIR) can be estimated as:

$$FPIR \approx \frac{Number\ of\ alerts - Number\ of\ Confirmed\ identifications}{Number\ of\ recognition\ opportunities}$$

With increasingly accurate facial recognition systems, sufficient recognition opportunities and transactions must be recorded in order to provide a measure FPIR (in the order of thousands or tens of thousands). It is therefore almost impracticable and very expensive to measure FPIR in a 'closed' trial using just volunteers.

5. Recommendations

- Prior to deployment of LFR capability, organizations should pay due regard to existing data, legal & ethical obligations
- Due diligence should be paid to baseline system algorithm accuracy and demographic differential performance
- The Concept of Operations should inform appropriate system deployment parameters
- Additional testing and system tuning should be undertaken that are aligned to the Concept of Operations and operational data
- Testing must conform to appropriate standards and guidelines (such as included in this document)
- LFR systems should be designed & deployed such that privacy by design is built in

- There must be an operator at the assessment and decision making stage (human-in-the-loop)
- Policies should be put in place around post deployment data retention and use

References

[1] Key considerations for responsible development and fielding of Artificial Intelligence, July 2020 https://www.nscai.gov/wp-content/uploads/2021/01/Key-Considerations-for-Responsible-Development-Fielding-of-AI.pdf

[2] Biometrics Institute Good Practice Framework https://www.biometricsinstitute.org/biometrics-institute-good-practice-framework/

[3] Surveillance Camera Commissioner, Facing the Camera https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/940386/6.7024_SCC_facial_recognition_report_v3_WEB.pdf

[4] Algorithms in Policing – Take ALGO-CARE http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/science-and-technology-committee/algorithms-in-decisionmaking/written/69002.html

[5] Face recognition Policy Development Template https://bja.ojp.gov/sites/g/files/xyckuh186/files/Publications/Face-recognition -Policy-Development-Template-508-compliant.pdf

[6] College of Policing Live Facial Recognition, Authorized professional practice https://www.college.police.uk/app/live-facial-recognition?s=

[7] Standard Operating Procedure (SOP) for the overt deployment of Live facial recognition (LFR) Technology https://www.met.police.uk/SysSiteAssets/media/downloads/force-content/met/advice/lfr/policy-documents/lfr-sop2.pdf

[8] Metropolitan Police Service Legal Mandate https://www.met.police.uk/SysSiteAssets/media/downloads/force-content/met/advice/lfr/new/lfr-legal-mandate-v.3.0-web.pdf

[9] NIST Face recognition Vendor Test (FRVT) https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt

[10] Face recognition Vendor Test (FRVT) part 3: Demographic Effects https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf

[11] Facial Recognition Technology in Law Enforcement Equitability study; Final Report, National Physical Laboratory 2022 https://science.police.uk/site/assets/files/3396/frt-equitability-study_mar2023.pdf

[12] ISO/IEC 30137-1:2019 Information technology — Use of biometrics in video surveillance systems — Part 1: System design and specification https://www.iso.org/standard/64935.html

[13] Face In Video Evaluation (FIVE) Face recognition of Non-Cooperative Subjects https://nvlpubs.nist.gov/nistpubs/ir/2017/NIST.IR.8173.pdf

[14] Home Office Scientific Development Branch, facial Image Test Targets https://www.gov.uk/guidance/cast-resources-for-the-crime-prevention-industry

[15] Standard Guide for Capturing facial Images for Use with facial recognition Systems https://fiswg.org/FISWG_Guide_for_Capturing_facial_Images_for_FR_Use_v2.0_20190510.pdf.

[16] Technical Report; Portrait quality reference facial images for MRTD https://www.icao.int/Security/FAL/TRIP/Documents/TR%20-%20Portrait%20Quality%20v1.0.pdf

[17] NISTIR 8271 Face recognition Vendor Test, Part 2: Identification https://pages.nist.gov/frvt/reports/1N/frvt_1N_report.pdf

[18] Phillips, P. , White, D. , O'Toole, A. and Norell, K. (2017), Human Factors in Forensic Face Identification, Biometrics in Forensic Sciences, Springer-Verlag GmbH, Heidelberg

[19] Phillips, P. , White, D. , O'Toole, A. , Hahn, C. and Hill, M. (2015), Perceptual expertise in forensic facial image comparison, Proceedings of the Royal Society B-Biological Sciences https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=917560

[20] Guide for Role-Based Training in facial Comparison https://fiswg.org/fiswg_guide_for_role-based_training_in_facial_comparison_v1.0_20200717.pdf

[21] Minimum training criteria for assessors using facial recognition systems https://www.fiswg.org/fiswg_min_training_criteria_for_assessors_using_fr_systems_v1.0_20200717.pdf

[22] ISO/IEC 19795-1:2021 'Information technology – Biometric performance testing and reporting – Part 1: Principles and framework https://www.iso.org/standard/73515.html

[23] ISO/IEC 30137-1:2019 Information technology — Use of biometrics in video surveillance systems — Part 2: Performance Testing & Reporting (in draft) https://standardsdevelopment.bsigroup.com/projects/2017-00948#/section