

Testing of Demographic effects in an Operational Live Facial Recognition Video System

Metropolitan Police Service - UK



**METROPOLITAN
POLICE**

Official

NEW
SCOTLAND
YARD

Outline

- Brief background to use of Live Facial Recognition (LFR) by the Met
- Why testing for demographic differential performance is important
- Demographic differential testing in an operational (LFR) context
- Test Strategy
 - Subjects
 - Performance Metrics
 - Environmental Factors
- Standards

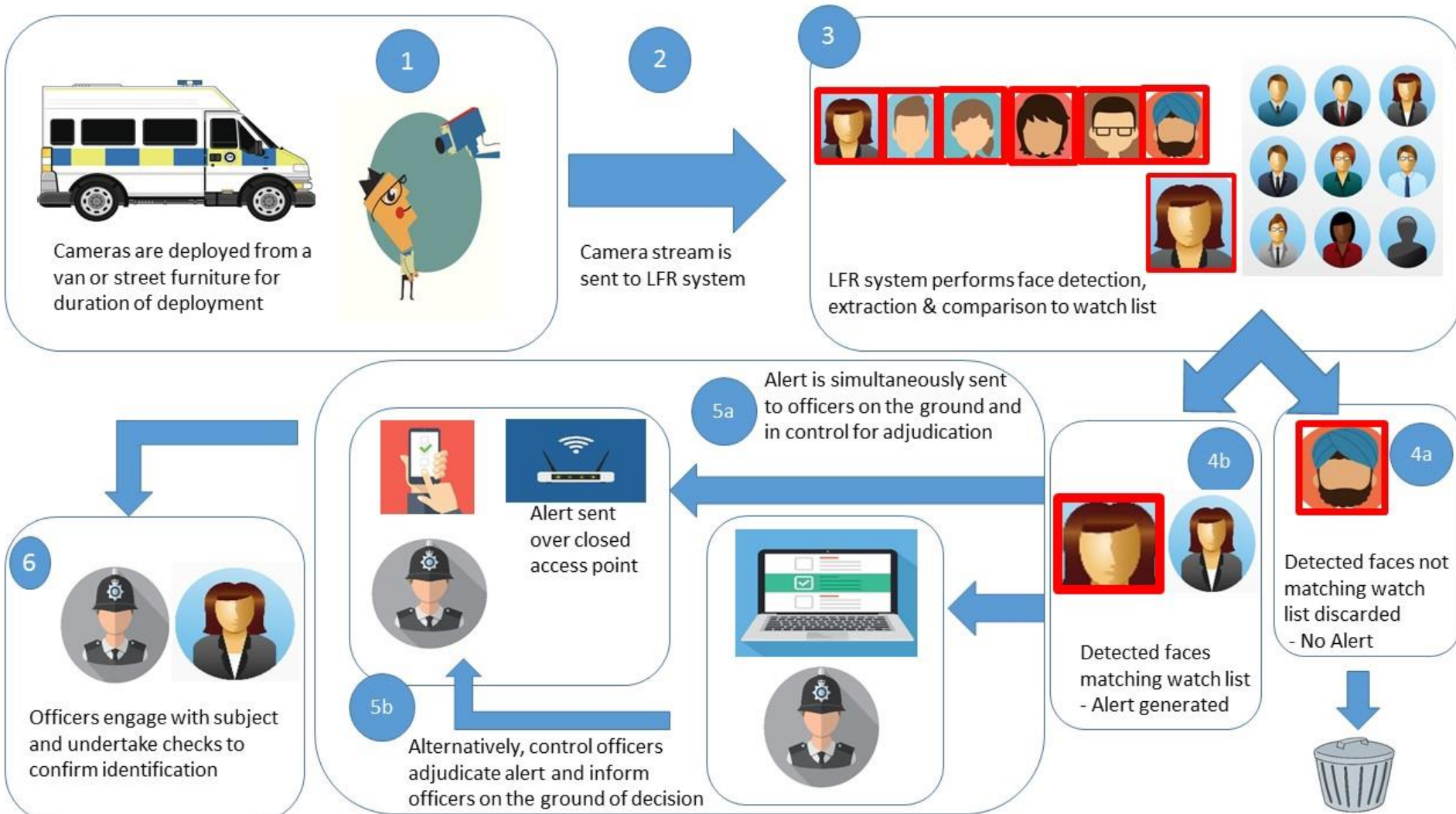
Brief Background of LFR by the Met



**METROPOLITAN
POLICE**

NEW
SCOTLAND
YARD

Brief Background of LFR by the Met



**METROPOLITAN
POLICE**



Brief Background of LFR by the Met

SUMMARY OF TACTICAL OUTCOMES	
Number of deployments	10
Combined duration of deployments	Approx. 69 hours
Watchlist size	Ranging from 42 to 2401
Recognition opportunities (number of people appearing video)	Approx. 180,000
Number of people engaged by a police officer following alert by the facial recognition system	27
Number of alerts confirmed correct at engagement	10
Actions / Arrests as result of alert	9

<https://www.met.police.uk/SysSiteAssets/media/downloads/central/advice/met/facial-recognition/met-evaluation-report.pdf>



Boston bans police from using facial recognition technology

NEWS

Court of Appeal warns of race bias in facial recognition technology

IBM Abandons Facial Recognition Products, Condemns Racially Biased Surveillance

June 9, 2020 · 8:04 PM ET

Amazon Suspends Police Use of Its Facial-Recognition Technology

Move comes after IBM said it was curtailing its facial-recognition activities amid widespread concerns about bias

**Big Brother is not only watching you—
he's identifying you.**



**METROPOLITAN
POLICE**

**NEW
SCOTLAND
YARD**

Testing for Demographic Differential Performance

Human
Rights
Act
1998

Article 8 - Privacy Considerations

- Engaged by searching using FR. This is because the search involves biometric processing of sensitive data – both those on a watchlist and the person in front of the camera
- All actions need to be necessary and proportionate.

Equality
Act
2010

Data Protection (DPA) – Law Enforcement Processing

- It needs to be 'strictly necessary' to search for anyone using FR.
- The MPS needs to understand statistical accuracy (as part of the accuracy principle) and bias (as part of the need to process data in a way that is lawful and fair).
- The ICO makes this clear and that vendor claims on these points should be challenged and verified.

Data
Protection
Act
2018

Public Sector Equality Duty (PSED)

- To use FR lawfully, the MPS will need to take reasonable steps to satisfy itself, either directly or by way of independent verification, that the algorithm does not have an unacceptable bias.
- This means quantifying the statistical accuracy and demographic performance of an algorithm



METROPOLITAN
POLICE

NEW
SCOTLAND
YARD

Testing in Operational Context

- A recent NIST report demonstrated a uniform distribution of scores across different demographics from the NEC-3 algorithm, however, the report clearly states that ***'it is incumbent upon the system owner to know their algorithm'***.
- It is not feasible to run a bench test to assess the demographic differential of the M30 algorithm in the context of Live Facial Recognition, as it is not possible to generate the number of biometric transactions required or the variability of conditions.
- The NIST report does not examine the effect of demographics on the acquisition of facial images, which is an important factor in operational live facial recognition.

Demographic Differential Test Strategy

The MPS will work in conjunction with the National Physical Laboratory to collect data over a series of operational deployments in order to run off line tests to measure demographic differential performance



Factors (and challenges) to Consider

- **Subjects**

 - Demographic make up of test subjects

 - Demographic make up of subjects on the Watchlist

 - Demographic make up of 'the crowd'

- **Performance metrics**

 - Failure to detect rate

 - False Positive Identification Rate

 - False Negative Identification Rate

- **Environment**

 - Do conditions (light/rain/other) impact on performance

Factors & Challenges to consider - Subjects

Demographic make up of test subjects

200 Subjects

50 IC1 Male

50 IC1 Female

50 IC3 Male

50 IC3 Female

Subject ID

Sex: Male/Female

IC Code: 1/3

Age

Height

Demographic make up of 'the crowd'

Not a trivial task

Will have to be estimated

Demographic make up of subjects on the Watchlist

Self defined ethnicity

May not easily fit into demographic test categories



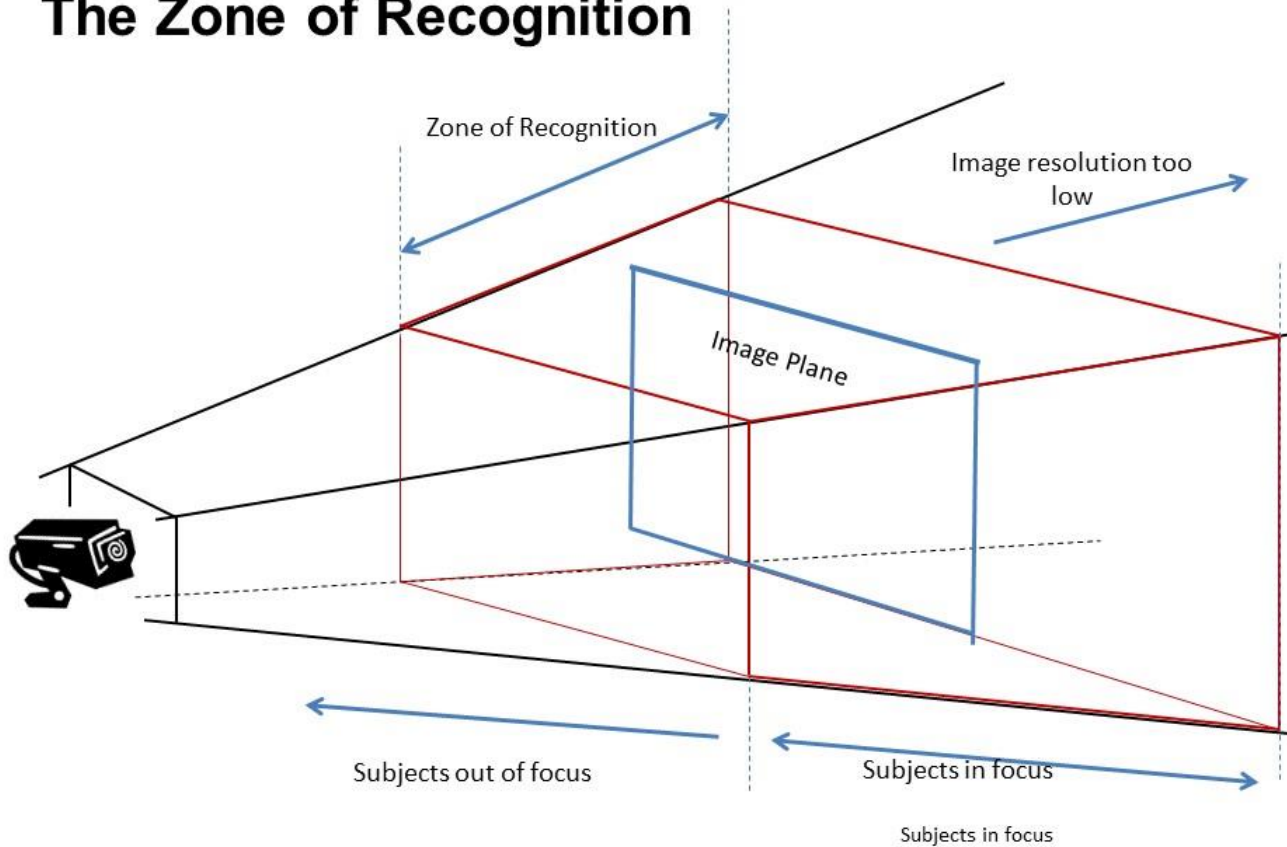
**METROPOLITAN
POLICE**

NEW
SCOTLAND
YARD

Factors & Challenges to consider - Performance

Failure To Detect Rate

The Zone of Recognition



**METROPOLITAN
POLICE**

NEW
SCOTLAND
YARD

Factors & Challenges to consider - Performance

False Positive Identification Rate

$$\text{FPIR}(N,T) = \frac{\text{Num. recognition opportunities of subjects not on the watchlist that generate an alert}}{\text{Num. recognition opportunities of subjects not on the watchlist}}$$

Need to aim for at least 30,000 recognition opportunities across the deployments

How do you measure the total number of recognition opportunities for subjects not on the watchlist?



Factors & Challenges to consider - Performance

False Negative Identification Rate

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

How do you Ground Truth the total number of recognition opportunities for subjects on the watchlist?



Factors & Challenges to consider - Environmental



**METROPOLITAN
POLICE**

NEW
SCOTLAND
YARD

Standards Documents

DIS ISO/IEC 19795-1:2020 Information technology – Biometrics performance testing and reporting – Part 1: Principles and Framework

ISO/IEC 19795-2:2007 Information technology - Biometric performance testing and reporting, Part 2: Testing methodologies for technology and scenario evaluation

ISO/IEC 19795-6: 2012 Information technology – Biometrics performance testing and reporting – Part 6: Testing methodologies for operational evaluation

ISO/IEC 30137-1: Information technology -- Use of biometrics in video surveillance systems – Part 1: System design and specification

CD ISO/IEC 30137-2: Information technology – Use of biometrics in video surveillance systems – Part 2: Performance testing and reporting

