Ethical issues arising from the police use of live facial recognition technology

Interim report of the Biometrics and Forensics Ethics Group Facial Recognition Working Group, February 2019

This briefing document outlines some of the ethical issues raised by the use of live (real-time) face recognition technology for policing purposes. It focuses on the use of this technology in relatively 'controlled' environments; namely public spaces where people are gathered and relatively static (for example, concert venues, sports stadiums, public rallies) and those with clearly defined entry and exit points or where people are 'channelled' past the cameras (for example, [approaches to] railway stations, airports, shopping centres, political marches or demonstrations).

The Biometrics and Forensics Ethics Group (BFEG) commissioned this report in response to the recent field trials of live facial recognition (LFR) undertaken by South Wales Police (SWP)[1] and the Metropolitan Police Service (MPS). This report outlines a framework of ethical principles that should be taken into consideration when developing policy on the use of LFR technology for policing purposes.

Details of the MPS and SWP trials have been covered in a series of reports published by:

- Big Brother Watch;[2]
- the London Policing Ethics Panel;[3] and
- the Universities' Police Science Institute and the Crime and Security Research Institute at Cardiff University.[4]

These outline the trial design, evaluate trial performance and raise a number of ethical and legal issues pertaining to these deployments. The aim of this report is to provide a short and general briefing that applies not only to the recent field trials, but also to the use of LFR in policing contexts more widely; it should be read in conjunction with the aforementioned reports.

Terminology and definitions

Police use of facial recognition technology has variously been described as 'live facial recognition', 'automated facial recognition' (by the Metropolitan Police Service/South Wales Police) and 'assisted facial recognition'.

Below the Biometrics and Forensics Ethics Group outlines a definition of live facial recognition technology based upon the International Standards Organisation (ISO) biometric vocabulary (ISO 2382-37) that is used throughout this document.

Biometric recognition is the automated recognition of individuals based on their biological and behavioural characteristics, for example, facial image, DNA, voice and gait.

Automated recognition implies that a machine-based system is used for the recognition, either for the entire process or assisted by a human being.

Live facial recognition (LFR) is the automated one-to-many 'matching' of near real-time video images of individuals with a curated 'watchlist' of facial images.

In the recent field trials LFR was used to **assist recognition** of persons of interest on the watchlist; this meant that police personnel were required to verify/override a possible match identified by the system (a system alert) and decide what actions, if any, should be taken on the ground.

During the evidence gathering process the BFEG facial recognition working party gathered evidence from representatives from:

- SWP and the MPS;
- academics from Cardiff (Martin Innes and Bethan Davies) and Essex (Peter Fussey) Universities, who have undertaken evaluations of the SWP and MPS field trials respectively;
- the Police Digital Service at the Home Office;
- the Defence Science and Technology Laboratory (DSTL);
- the Biometrics, Surveillance Camera and Information Commissioners' Offices;
- the Forensic Science Regulator; and
- civil society groups (Big Brother Watch and Liberty).

2. Technical issues

There is a substantial body of scientific research on the development of facial recognition algorithms, platforms, and systems that may be used for live facial recognition (LFR). This briefing document focuses on four issues that may affect the performance of LFR in policing contexts:

- data and training of the algorithms;
- the generation of outputs;
- the role of human operators; and
- deployments 'in the wild'.

Data and training LFR algorithms

Biometric technologies for facial recognition require machine-learning algorithms that have been trained on a dataset of labelled images.[5] The system can only 'recognise' faces within the parameters of the data that it has been trained on and previously exposed to. If certain types of faces (for example, Black, Asian and Ethnic Minority faces or female faces) are under-represented in LFR training datasets, then this bias will feed forward into the use of the technology by human operators. There have been high-profile scientific concerns that there is intrinsic potential racial and gender bias within LFR systems.[6]

Software and the generation of outputs

LFR is a technology that is probability based; it provides a probability of a 'match' between the captured image from the environment and an enrolled image on a watchlist. Multiple factors affect the probability of a match including:

- the quality of the enrolled images (pixel size, lighting, background and custody images versus social media, etc.);
- the quality of captured images;
- the algorithm's matching performance;
- the size of the watchlist;
- the environmental conditions (principally, but not limited to, lighting and camera position) where the image is captured;
- the thresholds that are set to determine a match on any biometric decision (determining the number of false and correct matches);
- whether a 'match' instigates a near real-time response or not; and
- whether the response includes a human who decides to take further action or to overrule the machine-generated biometric match.

The role of human operators

A key aspect of the use of this technology is the relationship between the output of the LFR and human operators' responses. The LFR software does not decide how the output will be interpreted and used/acted upon; this decision is the responsibility of the system operator. In the field trials police personnel were required to verify or override a possible match identified by the system and then decide what action should be taken (for example, intervention, identity verification, arrest).

A concern here is that an error, bias, or (in)accuracy in algorithmic output results in biased decision-making on the part of human operators. For example, if the system generates many 'correct' matches, then operators may start to defer to the algorithm's decision and act upon all matches without first verifying match accuracy.

Alternatively, if the system generates many false matches operators may begin to ignore or override all outputs, thereby missing correct matches. Finally, if the thresholds are set too high and too few matches are generated, operators may adjust the thresholds to produce more (potentially false) matches, which may result in more interventions with the attendant ethical consequences.

Deployments of LFR 'in the wild'

Where machine learning is taking place through the exposure of the algorithm to new sources of data in a public space, every police trial is potentially an operational deployment and every operational deployment is experimental and trial-like. This inherent ambiguity means that it is difficult to discern the purpose of the recent police field trials; were they police operations or experiments? This raises questions about:

- securing consent for 'trial' participation;
- the nature and composition of the watchlists (whether they should be simulated or contain images of persons of interest); and
- the extent to which field trials risk undermining public confidence and trust in policing.

3. Conclusions

There are a number of questions about:

- the accuracy of live facial recognition (LFR) technology;
- its potential for biased outputs and biased decision-making on the part of system operators; and
- an ambiguity about the nature of current deployments.

There is a need to differentiate the errors and biases that are inherent to the design and training of the technology from those that are introduced when a human operator decides on an action on the basis of the system output.

In addition, the Biometrics and Forensics Ethics Group (BFEG) notes the lack of independent oversight and governance of the use of LFR. Pending the development of a legislative framework the BFEG recommends that police trials of LFR should comply with the usual standards of experimental trials, including rigorous and ethical scientific design. The BFEG has drafted a number of ethical principles that can be used to inform these deployments and frame policy-making which can be found at Annex A. This is accompanied by a set of questions that arise when live facial recognition is used in policing contexts which can be found at Annex B.

References

1. SWP have trialled this technology in two modes AFR-IDENTIFY and AFR-LOCATE; only the AFR-LOCATE trials involve real-time one-to-many facial matching.
2. Big Brother Watch (2018) Face Off: the lawless growth of facial recognition in UK policing, May 2018.
3. London Policing Ethics Panel (2018) Interim Report on Live Facial Recognition, July 2018.
4. Davies, B., Innes, M. and Dawson, A. (2018) An Evaluation of South Wales Police's Use of Automated Facial Recognition, September 2018.
5. Parkhi, O. M., Vedaldi, A., Zisserman, A. (2015) Deep Face Recognition, Visual Geometry Group, University of Oxford.
6. Buolamwini, J. and Gebru, T. (2018) 'Gender shades: intersectional accuracy disparities in commercial gender classification', Proceedings of Machine Learning Research 91, pp 1–15.
7. Susskind, J. (2018) Future Politics: Living Together in a World Transformed by Tech. Oxford: Oxford University Press, p 282.
8. On biases in predicting recidivism see Angwin, J., Larson, J., Mattu, S. and Kirchner, L. (2016) 'Machine Bias', Propublica, May 2016.

Authors

This report was authored by the Facial Recognition Working Group of the Biometrics and Forensics Ethics Group. The members of the group were:

- **Professor Nina Hallowell (Chair), Oxford University;**
- **Professor Louise Amoore, Durham University;**
- **Professor Simon Caney, Warwick University; and**
- **Dr Peter Waggett, IBM**

Annex A

Ethical principles to inform the use of live facial recognition

The following ethical principles (based upon those developed by the Biometrics and Forensics Ethics Group) should be taken into account when considering the deployment of live facial recognition (LFR) or other automated biometric recognition technologies for policing purposes.

1. Public Interest. The use of this technology is permissible only when it is being employed in the public interest. In some cases, this might be straightforward, for example, there is a public interest in being able to identify those engaged in criminal activity. Other cases may be less straightforward.

2. Effectiveness. The use of this technology can be justified only if it is an effective tool for identifying people.

3. The Avoidance of Bias and Algorithmic Injustice. For the use of the technology to be legitimate it should not involve or exhibit undue bias. This can be unjust in two ways. First, some kinds of misrecognition are inherently demeaning and insulting.[7] Second, technology with these biases can result in unequal and discriminatory treatment of some individuals (for example, members of some groups may be much more likely to be detained and/or required to identify themselves). Automated biometric recognition systems (including data training sets) that will be used in public places should be to be open to scrutiny and effective oversight.

4. Impartiality and Deployment. If the technology is deployed for policing purposes it must be used in an even-handed way. For example, it should not be used in ways that disproportionally target certain events, but not others, without a compelling justification.

5. Necessity. Individuals normally have rights to conduct their lives without being monitored and scrutinized.

(a) Given that the use of the technology interferes with these rights, such technology can be used only if other, less invasive, techniques are not available.

(b) Furthermore, the technology should be used in ways that minimize interference with people engaging in lawful behaviour.

6. Proportionality. In addition to meeting a 'necessity' requirement, the technology should also meet a 'proportionality' requirement. That is, it can be permissible only if the benefits are *proportionate* to any loss of liberty and privacy. The benefits have to be sufficiently great so as to justify any interference with other rights.

7. Impartiality, Accountability, Oversight and the Construction of Watchlists.

(a) If humans (or algorithms) are involved in the construction of watchlists for use with the technology, it is essential that they be impartial and free from bias.[8]

(b) The construction of 'watchlists' needs to be subject to oversight by an independent body.

8. Public Trust. If the technology is to be used for policing purposes it is important that those using it (either in operational deployments or trials) engage in public consultation and provide the rationale for its use.

9. Cost-effectiveness. Any evaluation of the use of this technology needs to take into account whether any resources it requires could be better used elsewhere.

Annex B

Questions arising when live facial recognition is used in policing contexts

These questions to accompany the ethical principles are **not** to be treated as exhaustive, or as a checklist, but are intended to aid interpretation of the ethical principles when deploying live facial recognition (LFR).

1. Public Interest

- Why is LFR being deployed in this instance? (Crime prevention, intelligence gathering, etc.)

2. Effectiveness

- How accurate is this technology? (How are false positive/negative rates calculated?)
- Has the LFR technology been validated using ground truth datasets?
- What are the criteria for successful deployment? (True positive matches/no false positive matches, increased arrests, less criminal activity/ fewer arrests?)
- What is the quality of captured images?
- How is the system set up? (The importance of camera position and the network over which data are transmitted.)
- What is the trade-off between speed and accuracy? (System features.)
- How quickly can police officers respond to a match? (Location of the system in the field.)
- What information do field officers receive about the match? (Is the information detailed enough to inform an accurate identification and intervention?)
- What training do human operators have?
- Is human operator behaviour assessed/monitored for algorithmic deference/aversion?
- How is human operator error measured?

3. The Avoidance of Bias and Algorithmic Injustice

- Has algorithmic bias been taken into account?
- How is algorithmic bias measured?
- What is the nature of the data in the training datasets?

4. Impartiality and Deployment

- How are deployment sites decided?
- Who decides where LFR is deployed?
- Has a community impact assessment been undertaken?

5. Necessity

- What is the legal basis, if any, for the use of this technology?
- Does the watchlist include enrolled images of children?

6. Proportionality

- What is the purpose of deployment of LFR?
- Is the use of LFR proportionate?
- What are the costs (to individual liberty) and benefits (for public safety) of the use of LFR?
- Is retention of captured images or data proportionate?

7. Impartiality, Accountability, Oversight and the Construction of Watchlists

- Who has oversight of these deployments?
- How will the use of LFR be evaluated?

- Who compiled the watchlist?
- How big is the watchlist?
- Why this watchlist?
- Where are enrolled images on the watchlist derived from?
- How accurate are enrolled images on the watchlist?
- What guidelines have been used for the compilation of the watchlist?
- Who has oversight of the compilation of the watchlist?
- Are any captured images or data stored?
- How long are captured images or data retained after deployment?
- Where are captured images and data stored after deployment?
- Who has access to captured images or data?
- If captured images and data are shared with other organisations, who are they shared with and why?

8. Public Trust.

- Is LFR deployed in a trial or operational context?
- How extensively is the LFR deployment advertised in the community?
- How aware is the general public about this deployment?
- Is there adequate transparency about this deployment?
- Can members of the general public easily find out information about the deployment?
- If an oversight board has been set up is there public representation on this board?

9. Cost-effectiveness

- Is the use of LFR cost-effective?