The Future of Video Analytics for Surveillance and Its Ethical Implications

Andrew A. Adams[*] James M. Ferryman[**]

Keywords: Surveillance, CCTV, Video Analytics, Ethics, Regulation, Facial Recognition, Computer Vision, Cognitive Vision

Abstract

The current state of the art and direction of research in computer vision aimed at automating the analysis of CCTV images is presented. This includes low level identification of objects within the field of view of cameras, following those objects over time and between cameras, and the interpretation of those objects' appearance and movements with respect to models of behaviour (and therefore intentions inferred). The potential ethical problems (and some potential opportunities) such developments may pose if and when deployed in the real world are presented, and suggestions made as to the necessary new regulations which will be needed if such systems are not to further enhance the power of the surveillers against the surveilled.

[*] Centre for Business Information Ethics, Meiji University, Japan. This work was supported by the UK's EPSRC (EP/G069808/1) and Japan's JSPS (Kakenhi (B) 24330127).

[**] Computational Vision Group, School of Systems Engineering, University of Reading, UK.

This work was supported by the European Union project ARENA (FP7-SEC-2010-1: 261658)..

However, this paper does not necessarily represent the opinion of the European Community, and the European Community is not responsible for any use which may be made of its contents.

1 Introduction

The global video surveillance market is expected to reach £25 billion by 2015 (MarketResearch.com, 2011). The rapid proliferation in the number of CCTV installations worldwide, in areas such as shopping centres, underground stations, and airports, has expedited the demand for automatic methods of processing their output. This is because human operators cannot effectively monitor large numbers of cameras for long periods of time. (Tickner & Poulton, 1973; Donald, 2005). Naturally, this has generated significant interest in machine vision-based systems that can augment human operators. The systems are envisaged as providing intelligent filters directing the operators attention to cameras observing events that are of interest, as well as assisting the operators in tasks such as tracking individuals through a shopping centre. "Interesting" events may be ones defined by users as being in some way anomalous or which have safety, security or threat implications.

Some limited automated analysis of CCTV images has been deployed over wide areas, such as the Automated Number Plate Recognition (ANPR) systems in use to enforce congestion charging in London, average speed limits through roadworks, and generally throughout Britain's national road networks. Interest from law enforcement and private security in automated capabilities for CCTV systems, either to replace fallible and limited human operators, or to enhance both live and post-hoc investigative use of visual surveillance systems, is significant. Funding has been and continues to be available from national (and EU) funding bodies for both basic research in computer vision which addresses the fundamental capabilities needed for automated CCTV analysis and for translational research on integrating new algorithms and approaches into complete back-end system ready for integration into deployable CCTV operating systems.

In this paper, the current state of the art (in Section 2) in both new analysis algorithms and the directions of research in the field (in Sections 3 and 4) is presented. An overview of the state of regulation of CCTV is then given (in Section 5) before an analysis of the new ethical and regulatory threats and opportunities created by the likely success of current research is given (in Section 6).

In discussing this area, it is useful to first describe a typical surveillance system and to define a few terms of art in computer vision and CCTV systems, so that readers unfamiliar with the technical terminology are enabled to follow the description. Although some CCTV deployments involve a single camera, most involve multiple cameras, the aggregation of the field of view of all connected cameras forming the *surveillance area* of the system. The cameras are fed into a control room in real time and monitored by one or more operators. The images from each camera are usually archived allowing an operator to retrieve imagery (e.g. to provide evidence for a crminal trial) at a later time. One of the fundamental challenges of computer vision and other applications is to distinguish both in one frame from one camera, across a video sequence from one camera, and in comparing different views of the same area from more than one camera, what physical objects exist in the surveillance area and how they are moving (or not moving). These activities are referred to as *object detection* and *object tracking*. Once objects are detected and tracked, interpretation of their movements, particularly for objects identified as human beings, are subject to *behaviour analysis* based on various visual information (e.g. posture changes), movement within the surveillance area, and other information (time of day, entry/exit tracking for secure areas, learned or a-priori defined behavioural models etc.). The combination of high level multi-algorithm CCTV analysis with cognitive science techniques covering both human psychological states and artificial intelligence interpretations of data and models is referred to as *cognitive vision* as a research area. The overall capability to automatically analyse video images to extract objects, detect events, and to perform behavioural analysis, is referred to as *video analytics*.

2 State of the Art in Automated CCTV Analysis

Despite the need, automated CCTV analysis has been a challenging domain for computer vision over the last few decades. Despite significant advancements in cameras (BBC, 2012), processing and storage technologies, technological development in video analytics is progressing more slowly. Video analytic research includes detection, classification, tracking, and event and behavioural (including threat) recognition (Dee & Velastin, 2008).

The first stage of a surveillance system for people and/or object monitoring is detection and classification. The stage consists of localising new objects entering the surveillance area. In the case of people, this might be an individual, a group or a crowd. Indoor scenes tend to be simpler as ambient lighting is more controlled. Particular attention recently has been given to person detection with state-of-the-art methods based on Histogram-of-Gradients (HOG) feature in combination with other features, and classifier learning (Dollár *et al*., 2012). Random forests and ferns have now established themselves as a fast and reliable appearance descriptor for object classification (Evans *et al.,* 2012). Classifier grids, in which a separate classifier is trained for each image location, have also been shown to be a powerful method for object detection, however the approach is scene specific and therefore not practicable for imagery gathered from areas which change or from mobile cameras. In terms of face recognition researchers are largely operating on visible or near-infrared imagery. In the US National Institute of Standards and Technology's (NIST's) assessment of biometric face recognition in still images, the error rate halves every two years. In 2010, the best face recognition method matched 92% of mugshots to one out of 1.6m images (The Economist (2012)). Recent reviews of methods for object detection and tracking are given in Hu *et al.* (2004) and Yilmaz *et al.* (2006).

An understanding of the type of object identified can then allow the vision system to more accurately track the object through the scene. The process usually takes account of the likely object dynamics and expected changes (or constancy) in shape and appearance over time. State-of-the-art tracking systems are capable of tracking several persons or vehicles simultaneously through short term occlusions and over short distances (i.e. within a surveillance area where camera views overlap each other so that line of sight is only ever briefly lost). The success of such approaches relies on a combined robust object representational and tracking framework that aggregates all available evidence over time. One effective object representation is based on a set of features and includes methods based, for example, on local point features or region-based statistics. Probabilistic tracking methods have proven to be especially effective in noisy conditions (e.g. where illumination changes make accurate extraction of object silhouettes via image analysis difficult) for tracking of people (Ben Shitrit *et al.* (2011)), in which the framework enables explicit fusion of information from multiple cameras. Research has also been conducted into tracking groups which, in addition to tracking individual members of a group, model situations in which groups of people split and merge (Zaidenberg *et al.* (2012). When objects are apparently absent (i.e. not observed) within the surveillance area for longer intervals, there are algorithms which can be employed to attempt to re-identify lost objects. All these methods rely on strong appearance models, operate over small areas, have limited applicability to dense, crowded environments, and do not generally operate in real time. They are thus primarily useful for after-the-fact investigation of activity within the surveillance area.

The final stage of a surveillance system is to attempt to understand the behaviours of tracked objects and the interactions between objects, and with the environment. A common approach is to apply a set of predetermined spatio-temporal rules that are correlated to what operators would interpret as “interesting”. A piece of luggage having been abandoned or a person moving in a restricted access area are typical examples of primitive events. The approach may be formalised through a hierarchical classification of these events and related scenarios (spatiotemporal combination of events). Such knowledge may then be modelled as an ontology to which it is possible to associate an inference engine to detect events automatically. The alternative approach to recognising events and behaviours is to infer them directly from representative data and validate the models against information extracted from the CCTV imagery. Many approaches are based on the concept of building a model of what is considered “normal” behaviour and determining anomalous behaviours by determining the degree of fit with the model. A surveillance focused overview of behaviour analysis methods may be found in Ko (2011).

The current generation of surveillance systems tend to exhibit their best performance in constrained scenarios where substantial contextual information can be exploited. They tend to perform with decreased performance when the monitored scene is unconstrained and/or complex in terms of the number of objects, their proximity to each other, their dynamics, and the amount of clutter in the scene (Dee & Velastin, 2008)). There may also be significant illumination variations between images of the same scene, for example due to environmental changes (lighting, shadows, weather). Image digitisation issues and poor resolution further hamper information extraction. All of these issues affect the performance of the surveillance system to some extent, potentially very significantly. Recent and future challenges in automated surveillance are reviewed in Gong *et al.* (2011); Coetzer *et al.* (2011); Dee & Velastin (2008); Dick & Brooks (2003).

3 Proven Prototypes for New Automated CCTV Capabilities

In recent years there have been a number of significant surveillance projects focussing on automated CCTV surveillance for safety, security or threat assessment applications. The UK EPSRC-funded REASON project (www.reason-cctv.org) developed a methodology to address robustness issues in monitoring and understanding people in public places. Given the needs of surveillance operators, the complexity of scenes under surveillance and limitations in current video analytics, next generation surveillance demands distributed, network infrastructures, greater robustness and adaptation in their video analytic algorithms and real-time operation. To address these needs a real-time distributed and scalable architecture was built and demonstrated through an implemented prototype (Valera *et al.* (2011)). The prototype included an approach to real-time robust segmentation of people within the monitored surveillance scene (Tweed & Ferryman, 2008) and human action modelling and recognition based on image-based extraction of silhouettes of people (Ragheb *et al.*, 2008; Orrite *et al.*, 2008). The project was focused on volume crime including vandalism, theft from vehicles in car parks and muggings, and had a strong emphasis on performance evaluation. Social, legal and ethical studies were performed as part of this project which, along with other projects, were the basis of the recommendations for a new legal viewpoint on CCTV video data presented by Adams (2007).

The EU-funded ISCAPS (Integrated Surveillance of Crowded Areas for Public Security) project (www.iscaps.net) had the general objective of reducing the risks of malicious events by providing efficient, real-time, user-friendly, highly automated surveillance of crowded areas. The scenarios considered included detection of abandoned luggage (with a formal definition derived for abandonment of luggage), collection of new video datasets of representative abandonment scenarios and performance evaluation of a range of detection methods on the collected benchmark data (PETS2006, 2006), detection of falling people (within, for example, a smoke filled metro station) using thermal imaging (Pham e*t al.*, 2007), and detection of anomalous behaviour involving people and vehicles. A significant number of the project requirements were met with the integration of existing technologies. However, a significant amount of technology development was required (i.e. video analytics, communications and system configuration) in order to improve the surveillance capabilities for the scenarios considered. In particular, the project supported the integration of a combination of sensors including video and thermal cameras, biometrics and optical authentication tags in order to improve the quality of identification and tracking of persons, goods and vehicles.

In the following sections two specific recent developments in abandoned bag detection and face image enhancement are considered in more detail.

3.1 Robust Abandoned Bag Detection

SUBITO (www.subito-project.eu) is a recently completed EU project aimed at developing a surveillance system for robustly detecting abandoned bags in public spaces and to identify and track the owner. The project built upon the set of spatiotemporal (abandonment) rules developed within the ISCAPS (www.iscaps.net; PETS2006, 2006) project (which related any individual near an identified bag) by introducing the concept of ownership and knowledge of social relations. The rationale was that a global analysis of the situation, rather than just examining each agent's behaviour independently, would be beneficial in many situations. The motivation for this is illustrated by a scenario similar to that of 2007 IEEE Performance Evaluation of Tracking and Surveillance (PETS) Workshop PETS2007 (2007) where a family or a group of friends comes together and one of them leaves his/her bag with the others. Any threat detection system treating the individuals independently would inevitably report incorrectly an abandoned bag, as the criteria specified in PETS2006 (2006) (that the bag is abandoned if the owner is further than *a* metres for more than *b* seconds) is fulfilled. For dealing with these more complex scenarios, it became necessary to derive a more complete activity analysis and the concept of social groups was introduced.

In SUBITO, an algorithm was developed, by project partner the University of Leeds, for automatic detection of social groups within crowds, based on the analysis of the way social relations influence the walking behaviour of group members. The method is based on the Social Force Model (SFM) of Helbing & Molnar (1995) and Moussaid *et al.* (2010), widely used in the crowd simulation community. In this, each individual's movement is influenced by notional forces operating between individuals. Depending on whether two individuals (a) know each other or (b) do not know each other, the SFM produces different sets of trajectories for these individuals. The method employed in this work (Sochman & Hogg, 2011) solves the inverse problem: knowing the trajectories (obtained from a Multi-Hypothesis Tracker), what are the social forces, and thus the relations, that caused that behaviour. The method is used in the SUBITO system to infer the social relations between the individuals in a scene and thereby to inform threat assessment. A rule-based system is used and the bag distance is checked against either an individual or the relevant group as derived using the Social Force Model. Once an alarm is raised, a Human Machine Interface interface provides the end user with the means to perform rapid browsing of both recorded and live video CCTV streams to identify the whereabout of the bag owner(s). The overall system operates in quasi-real time (typically a few seconds before an alert is generated.) In trials of the system using multi-camera surveillance of an outdoor scene involving staged abandoned bag scenarios with multiple persons, the system was able to correctly alert in all but one situation. The error was due to an incorrect assignment of two individuals to the same group because of insufficient evidence of their relationship prior to the abandonment. In comparison, a baseline implementation of the system, which does not exploit understanding of social relations, resulted in multiple failures in situations where individuals moved close enough to a bag preventing it being classified as unattended.

3.2 Face Identification: Hallucination and Superresolution

Significant issues with the capture of facial images from CCTV, particularly for evidential purposes, include low resolution and degradation due to occlusion and noise. There has been a significant amount of image analysis research work undertaken to enhance such imagery to permit more accurate identification. A major area of research is the "superresolution" approaches which is the process of recovering a high resolution image of a face, through combination of a set of multiple low-resolution images of the same face. For example, a number of snapshots of the same person may be captured as they move across the field of view of a single CCTV camera. The methods operate directly on the image data and exploit the spatiotemporal regularities between the images. Another approach is the hallucination-based method: reconstruction via inference of a high-resolution face image from a single low resolution image with the aid of a large library of other high resolution face images (Liu *et al.,* 2007). Park & Lee's (2008) approach is based on finding a linear model to learn the relationship between the set of high resolution faces and their smoothed and downsampled lower resolution versions, for example holistic-based on Principal Component Analysis (PCA) or support vector learning (Lee *et al.,* 2006). Some methods go one step further to improve the result through, for example, application of a Markov network to capture finer detail exhibited by the high frequencies. Lee *et al.* (2006) observe that "the experimental results [in their work] showed that the facial images synthesised ... were quite similar to the high-resolution, thus making it possible to recognise a person from a low-resolution image." It should be noted, however, that the problem by definition is ill-posed (as it is formulated as an inverse problem of estimating a high resolution image from low resolution image or images) and visual artifacts may be introduced as a result of the reconstruction process which is essentially "inventing information". Most of the research has also focused on facial images captured under constrained conditions (i.e. fixed illumination and pose). However, some research has been conducted which has examined multiple viewpoints (Jia & Gong, 2008).

4 Research Directions in Automated CCTV Analysis

While there have been significant advancements in computer vision with related impact in a significant number of application fields there remain several complex challenges in research terms. Furthermore, improvements in cameras, networking, storage and processing hardware have opened up new subdomains including the research and development of cooperative multi camera networks (Section 4.1), analysis of crowded scenes (Section 4.2), advanced behavioural understanding (especially recognition of intent – Section 4.3) and cognitive inspired capabilities (Section 4.4). Such research domains require breakthroughs in algorithm quality (including accuracy and robustness), not just increased raw power and transmission capability. This is especially true when considering the combinatorial explosion potential in analysing larger public spaces utilising tens, hundreds, or sometimes thousands of sensors.

4.1 Cooperative Multi Camera Networks

A robust visual surveillance system must be capable of tracking objects across a surveillance area past occurrences of partial or full temporary occlusion. Furthermore, tracked objects may disappear completely from view and re-appear some time later. The problem can be alleviated, to some extent, by employing a camera network (or networks) which contain multiple cameras occasionally with multiple overlapping fields of view. However, many pre-existing installed CCTV networks contain camera views with limited or no overlap. To address this issue, researchers have begun to seek methods which can associate and track people (the “person re-identification” problem) and other objects over extended areas, with missing information, over extended periods of time, and in an efficient manner (Bazzani *et al.*,2012)). Ideally, the relative views (determined by camera position, orientation, field of view, and occluders) of the cameras in a network, or cameras from different networks, are known. In practice, these things are often not measured, and cameras might also have moved (especially outdoors) after such measurements, which naturally makes multi-camera tracking much harder. Methods have been developed for calculating camera configurations from their visual output, compared against a known geographical location and/or views from cameras with overlapping fields of view.

4.2 Crowd Image Analysis

The analysis of crowded scenes is of great interest in a large number of applications including crowd management (e.g. development of strategies to ensure public safety) and footfall estimation (measurement of number of people present at an event or other public space.) Methods can be categorised as low-level: crowd person count and density estimation; mid-level: tracking of an individual or individuals within a crowd; or high-level: detection of separate flows and specific crowd events (PETS2009, 2009). Depending upon the number of people within the scene, methods are generally either focused on detecting and tracking of individuals, or for particularly dense scenes where tracking individuals as individuals is difficult, holistic approaches treat the crowd as a single entity moving along a given path or direction. For example, motivated by fluid dynamics, (Mehran *et al.*, 2010) applied the concept of streak flow which encapsulates motion information of the flow over a time interval. Recent surveys of crowd analysis and the application of computer vision techniques, are presented in Cezar *et al.* (2010) and Zhan *et al.* (2008). Given the importance of crowd analysis, the theme of the 2009 PETS Workshop (www.pets2009.net) was multi-sensor tracking and event recognition in crowded public areas. The datasets produced for that workshop were re-used in four subsequent workshops (including PETS 2010 (www.pets2010.net) and PETS2012 (www.pets2012.net)) allowing a comparative evaluation of a number of state of the art crowd image analysis techniques. While there are some very promising approaches there is still much further research needed to improve overall accuracy and robustness (Dee & Hogg, 2004; Dee & Hogg, 2006).

4.3 Cognitive Vision

As stated in Section 2 above, current automated surveillance systems operate in restricted domains in which algorithms operate well. However, the overall vision is to develop systems which can respond to and act in the real world. Cognitive systems need to acquire information through learning or association. The fundamental mode of operation for learning is that action precedes perception. This implies that cognitive systems must be developed in a full perception-action feedback cycle. AI-oriented and systems-oriented areas of cognitive computer vision include: knowledge representation, learning, reasoning about events and about structures, recognition and categorisation, and goal specification and achievement. The European Commission have identified the objectives of cognitive vision to include the development of robust cognitive vision systems acquiring and using knowledge for decision making, to focus on adaptive systems for real-time platforms and vision architectures permitting the development of novel computational frameworks, integrating multiple cues for scene modelling, recognising large number of objects and achieving cognition such as temporal learning and incremental learning. Based on this, an increasing number of projects have been undertaken which focus on the development and integration of cognitive components into surveillance systems. This includes the recently completed Co-Friend project (www.co-friend.net) whose aim was to design a framework for understanding activities taking place on an airport apron. A heterogeneous camera network, composed of wide angle and Pan/Tilt/Zoom cameras was deployed at Toulouse airport in France. Multi-camera data fusion, video analytics and feedback was exploited to achieve real time recognition of activities, for example, aircraft refuelling and baggage loading/unloading operations (Patino *et al.* (2010)). The scene understanding was assisted by machine learning, providing advanced reasoning capabilities compliant in a largely unsupervised way to variations and novelties. Other initiatives include the EUCogIII project (2012-) (www.eucognition.org) a European network for researchers in artificial cognitive systems and related areas who want to connect to other researchers and reflect on the challenges and aims of the discipline.

4.4 Recognition of Intent

The SUBITO project, described in Section 3.1 above, has advanced the state of the art in situational awareness and event detection through automatic understanding of social groups. This research has enabled a reduction in the number of false alarms for abandoned bag detection. However, if the number of false alarms are to be reduced even further then an even more complete understanding of situations involving people and the environment needs to be developed. This may include automatic inference of where a person is moving to in a scene. Taking the SUBITO abandoned bag scenario as an example, once the bag has been deposited it is important to ascertain whether the owner has simply moved away, for example, to purchase a ticket with the full intention to return, or has walked towards the exit. Research undertaken by the University of Leeds, as part of SUBITO, has studied this exact issue, one of intent modelling from visual cues. The approach is based on psychological studies which show that human intent can be inferred from incomplete visual observations. Leeds have developed a novel method for making a running prediction of the current destination for each individual tracked within a scene that can deal with indirect paths and random perturbations from an ideal path. Prior work included the development of two systems based on intentionality analysis within behaviour modeling. The first system includes the use of an ad-hoc model of goal-directed behaviour and recalculation of possible intentions of each individual at each frame (Dee & Hogg, 2004), and a second system employs models of navigation from the psychological literature (Dee & Hogg, 2006).

4.5 Interrogating Large Volumes of CCTV

There is a responsibility on government agencies and police forces worldwide to capture and review massive volumes of CCTV images during investigations. The advent of large numbers of CCTV cameras, of increasing quality in terms of resolution and frame rate, has led to an explosion in the amount of data that is gathered by the intelligence and law enforcement community. Taking the London bombings of 7th July 2005 as an example of a major investigation undertaken by the UK Metropolitan Police Service where CCTV footage was vital in understanding the sequence of events, 90,000 hard drives and video tapes from CCTV systems were seized totalling more than 6,000 hours of CCTV footage. The analysis of large volumes of CCTV, however, is a difficult and time consuming process because of both interoperability problems between the various surveillance systems on the market and the lack of application of video analytics. There is significant scope to apply video search and data mining methodology to aid in forensic video archive search.

4.6 Privacy Enhancing Technologies

The tracking technology described does not have to be used exclusively for video surveillance tasks *per se*. Instead some recent developments in image and video processing have placed emphasis on improving the accuracy of detection and tracking methods with the specific aim of increasing the privacy of individuals and vehicles under CCTV surveillance (Spindler *et al.*, 2008). The process of protecting privacy through processing of recorded video has been utilised for decades in television, for example by pixelisation or blurring the faces of witnesses or other anonymous persons, or the number plates of vehicles of dignitaries. A technique to automatically perform such blurring or masking of faces in video (based on colour discrimination) was patented by Sony in 2000 (Berger (2000)). Automatic masking of people in surveillance video was demonstrated in Senior *et al.* (2005) as a component of IBM's PeopleVision (www.research.ibm.com/peoplevision/videoprivacy.html). The company MindMancer has developed and installed systems with built-in anonymisation of people and/or "virtual walls", beyond which people are removed from the imagery. Privacy enhancing methods can be either irreversible (i.e. the original content is distorted and cannot be recovered) or reversible. Irreversible methods include applying conventional pixelisation and blurring methods at image level, and MPEG-7 video streams which only transmit information required for a monitoring system to make a decision without providing any other information (Annesley *et al.*, 2007)). Reversible approaches include reversible encryption or scrambling of visual information (still image or video) in either uncompressed or compressed frames (Dufaux & Ebrahimi, 2008) Examples include reversible scrambling of JPEG2000, MPEG-4 and H.264/AVC compressed content. Despite these developments, such privacy enhancing technologies and systems have not hit the market with overwhelming force mainly because neither the market nor the technology is mature enough. In particular, systems able to mask people or vehicle plate details must be able to accurately detect and track (and segment) people and vehicles in video streams, which remains a very challenging task in automated surveillance research. Most recently, the EU network of excellence VideoSENSE (www.videosense.eu) (Virtual Centre of Excellence for Ethically-guided and Privacy-respecting Video Analytics in Security) has been instigated to foster significant advances in the domain of ethically-aware data and video analytics.

5 The State of CCTV Ethics and Regulation

As discussed in (Adams, 2007)), the regulation of CCTV has been quite limited and primarily based on existing data protection rules designed to deal with more structured data. This approach led to the bizarre situation of video footage being personal data about everyone in view in some EU countries but only about a person being specifically targeted for surveillance in the UK. Hempel & Töpfer (2004) report on the inclusion in the German federal data protection law of explicit sections on CCTV, while in Denmark there is a ban on CCTV in public places.

The UK's coalition government elected in 2010 had promises to roll back the surveillance state (instituted by successive preceding administrations over at least twenty years (Norris & Armstrong, 1999)) in the manifestos of both parties. Although they have been criticised for backsliding on a number of key promises (Rowlands, 2011) the Protection of Freedoms Bill, going through the final stages of parliamentary process at the time of writing, does include a specific statutory basis for most CCTV operations, the creation of a Surveillance Camera Commissioner and a legally binding code of conduct for CCTV operations. Previously CCTV operation in the UK was covered by a code of conduct issued by the Information Commissioner's Office. The status of such a code in court was very limited, to the status of at most "best practice". While the new code proposed in the bill would not be made directly legally applicable[1] it is specified in the bill that the code is admissible in evidence and that courts may take into account failures to follow the code.

In Japan, CCTV is regulated only under the anyway very weak data protection legislation (Adams *et al.*, 2010). While security system installation and operation service providers advise clients to follow local city guidelines on CCTV usage within their premises (Adams *et al.*, 2012) such guidelines are only designed to cover the CCTV operations of the authorities themselves. Meanwhile, CCTV deployments in Japan are significant, in both public and semi-public areas and in particular there is large growth in

[1] i.e. violations would not automatically be grounds for a criminal or civil conviction

deployments in semi-private[2] areas such as company back offices. Reports from the security industry (Adams *et al.*, 2012) suggest that this is regarded as unproblematic and effectively unregulated.

Given the lack of clear regulations in many countries on current "dumb" CCTV systems, despite their broad installation and not only potential but demonstrated capacity for misuse (BBC, 2006), the development of the "smart" CCTV systems described above raise significant concerns.

6 Ethical Gazing into the Crystal Ball

As the capabilities of CCTV surveillance improve, so the pressure to deploy it will only grow. As Norris & Armstrong (1999) pointed out, political selling of CCTV as the solution to the (possibly politically created) fear of crime led to demands from the UK's populace for the installation of CCTV without significant prior evaluation of its efficacy and without plans to evaluate its actual impact. Sometimes sense prevails, such as the case of automated facial recognition systems installed but then abandoned in Tampa Bay and at Palm Beach airport due to their lack of utility (both false positive and false negatives rates), as reported by Introna & Wood (2004). The lack of hard evidence of the impact of CCTV deployment on the problems it supposedly solves or mitigates provided useful political cover since it can always be claimed that without the system the situation would be worse even if no improvement has been seen. Since the criticisms of Norris & Armstrong (1999) there have been modest efforts at evaluation of CCTV schemes. Welsh and Farrington (2009) reported on a meta-analysis of only 22 rigorous studies, for which the only statistically significant crime reduction which could be claimed was in deployments within car parks. Small crime reduction effects in other areas could be identified, but without sufficient statistical weight.

Recent EU-funded security research projects which include or are focused on CCTV systems development, such as SUBITO, IMSK (www.imsk.eu) and others, have been required by the EC to include ethical oversight, not just of their research efforts as was often the case in the past, but of the intended outcomes of the projects. Much of the focus of the ethical oversight in these cases has been on including a Privacy Impact Assessment process during the system developments, seeking out options for better privacy as part of the system design. Additionally, guidelines for those deploying the eventual systems are supposed to be produced, giving details of the privacy and other ethical risks that the systems can raise in use and providing advice on how to minimise their impact on privacy and potentially negative consequences of their use. These goals are worthwhile, although the success of their execution remains to be demonstrated, particularly when the

[2] Public, semi-public, semi-private and private spaces are defined in Adams (2007).

commercial partners engaged in these projects come to the point of selling the resulting product to a budget-conscious deployer who may be wary of the added costs of considering the ethical implications of their deployment and eager to simply get on and use it. This is where the EC and other governments need to come back into the frame again and demand such ethical oversight be embedded in considerations of whether, where and how to deploy such systems.

In the rest of this section, the ethical problems posed by the new capabilities under investigation or development, as detailed above, are considered.

6.1 Tracking, Privacy and Oppression

Tracking and identification of individuals and objects within a surveillance area has many uses, both legitimate and illegitimate. The introduction of CCTV cameras originally created a debate about the issue of whether remote viewing by camera was any different to local viewing in person, in places where the expectation or even the strong possibility of being seen existed. There are those who claim that CCTV as it exists at present is no problem in public or semi-public places and this was the constitutional interpretation in many places as described in Goold (2002, p.21 and p.26, [Note 6]):

> Despite declaring in Katz v. United States that the Fourth Amendment "protects people not places," since the late 1960s the Supreme Court has been highly resistant to the idea that privacy rights can extend to streets or other public areas.
>
> At present, this situation is similar to that under the European Convention on Human Rights. Although the Convention recognizes a citizen's right to "respect for his private and family life, his home and his correspondence," based on the decision of the European Court of Human Rights in Friedl v. Austria, it is unclear as to whether this right gives rise to any expectation of privacy in public places. See Friedl v. Austria, 21 European Human Rights Reports 83 (European Court of Human Rights, Jan. 31, 1995).

and that provided that notification is given even in semi-private spaces visual surveillance is acceptable (arguments for and against this are discussed by Introna (2000)). There are those, however, who point out the lack of reciprocality of gaze (there is no way of knowing who is on the other side of the camera whereas, usually, local seeing involves being seeable in return) such as Goold (2002) and the possibility of recording of exact images of activity (rather than simply relying on the memory of those present) which makes CCTV rather different to local direct surveillance such as Armstrong and Norris (1999, p.18). Similar arguments must now come into play as the automated analysis capabilities described above become deployable and deployed. At present one of the arguments for public space CCTV not being a significant invasion of privacy is the amount of resources needed to track individuals over time and through multiple systems is immense (see Section 4.5). Automated systems are moving in the direction that first any small number of selected individuals may be tracked, and eventually tracking of everyone within view in real time and associating that with prior activity may well become feasible.

This also raises significant questions regarding the security of video feed data. The present limits of what can be done with a video feed make interception by unauthorised parties not worthwhile in almost all cases. As the technology to make use of such footage improves, the value of such interception rises, and the dangers to privacy increase.

Tracking can be used in oppressive policing in both oppressive and democratic states. Consider the criticisms of the UK's Metropolitan Police regarding their handling of the G-20 Summit protests (Lyall, 2009). The reports of police obstruction into allegations of significant misconduct by officers and the use of crowdsourced video reported by Lewis (2009) show, however, that video can be used as powerful tools by both sides. In regulating access to both video footage and automatic analysis tools, the utility of such needs careful consideration in terms of both prosecution and defense, and in holding powerful groups to account as much as in maintaining order, as suggested by Brin (1998). Individuals' mobile phone photographs or videos raise questions on chains of evidence and reliability, of course as discussed by Coudert *et al.* (2011). As discussed elsewhere in this article, the reliability of crowdsourced images, which are frequently of shaky quality to begin with, may be further undermined by the possibilities of tampering. Where crowdsourced video footage is presented in support of eyewitness accounts (and in particular offers support for one side of differing accounts), such as in the case of Ian Tomlinson during the G-20 Summit protests (Lyall, 2009) then their provenance may well be less heavily questioned than where video footage is the only or primary evidence available.

6.2 Reconstruction of Facial Images

The concept behind this technology has been shown in movies and TV for many years, massively overstating the capabilities of most CCTV systems currently in practice. The concept that a series of low resolution still images can be parsed to extract the maximal information present in each image to develop a higher resolution version is sound in principle. The human mind does something similar, interpolating a sequence of television images from a DVD, say, into a higher resolution moving picture than each still image presents. Such processing is somewhat speculative in result, much more so than unprocessed images, which still require treating with caution when used for identification and particularly as evidence in criminal trials.

However, the recent developments by Liu *et al.* (2007) and Park & Lee (2008) of systems allowing the reconstruction of high resolution facial images based on single low resolution frames present much greater ethical concerns. These systems are using average facial image data, perhaps enhanced by cues such as skin tone or hair colour, to create an approximation of a face that could have resulted in the low resolution original image.

As shown by the tragic case of Jean Charles de Menezes (Independent Police Complaints Commission (UK), 2007) mistaken identity in security situations can cause irretrievable tragedy. Even in cases well below this extreme, mistaken identity can lead to severe disruption of life, ranging from the late Senator Edward Kennedy's inclusion on the no-fly list in the US (Florence, 2006) preventing his air travel for several days, to the arrest at gunpoint and holding in a police cell in Durban, South Africa of British pensioner Derek Bond on the request of the FBI (BBC, 2003).

Although reconstructed high resolution face data from low resolution images is unlikely to be acceptable as evidence in court, particularly in a criminal court, lack of clear regulations against its use in police investigations and private security operations is likely to lead to significant problems for completely innocent parties. The "reversion to the mean" inherent in the process increase the chances of a close random match between the generated image and a person who might have difficulty establishing their innocence. Existing reconstructive approaches such as sketch artists and photo-fit techniques combined with eyewitness identification have led to miscarriages of justice in the past and this new technique is at least as risky, if not potentially more so.

6.3 Cognitive Inference

The inference by automated systems of the state of mind of an individual, their likely intentions or even their previous behaviour, based upon external clues of posture, small scale body movements (hands touching the face, for example) or large scale movements (the pattern of movement within a surveillance area) has similar possibilities for over-reaction by security and law enforcement personnel. In their eagerness to prevent another public transport bombing, UK police officers made a series of errors in the de Menezes case. As the Independent Police Complaints Commission (UK) (2007) made clear, many though not all of these individual decisions were judgement calls made during a heightened state of alert. The algorithmic identification of individuals who pose a threat raise similar questions. As Reeves & Nass (2002) demonstrated, humans are generally unable to distinguish information from computer systems from information from other humans at a subconscious level. The presentation of the results from such cognitive interpretation systems therefore requires the highest level of consideration, again in order to prevent, in extreme cases, potentially lethal responses by law enforcement against innocent individuals. In particular the models used by such systems must be shown to be robust against false positives arising from individuals with safe but abnormal states of mind, such as those suffering from obsessive compulsive disorders.

7 Conclusions

Although deployments of automated CCTV surveillance to date have been quite limited, many recent research developments have been targeted at integrating prior work into robust mechanisms for real world applications and hence it is likely that deployment of such systems is no more than a few years away. The technical capabilities of even these developments are a further quantum jump from manual CCTV systems, just as those were over in-person surveillance. As such, significant attention needs to be paid to the ethical issues that are raised by these new technologies and appropriate regulations about their deployment and use put in place *before* they become the *fait accompli* of infrastructure.

The limitations of the systems need to be well-understood and widely published. Not only might this aid in the acceptance of such systems being deployed, but would also ensure that both operators and those subject to the surveillance have a decent understanding of the capabilities of such systems, rather than relying on either Hollywood representations or on the sales pitch of security systems companies. Such over-reliance is a significant worry in the deployment of systems such as face reconstruction and behavioural interpretation. Without clear guidelines on the limits of such systems the chances of over-reaction by security personnel to false positives could create serious consequences for the individuals falsely targeted.

In addition, the deployment of advanced CCTV systems should not be taken as a guarantee of perfect security. All security systems have vulnerabilities and these systems themselves will be vulnerable to attack in a number of ways, both directly through interception and possible replacement of signals or simply interrupting power or communications signals, up to attacks based on bypassing the trigger points in the analysis.

The further transfer of power from the surveilled to the surveillers, which such systems could represent, creates a broader problem. As discussed above, the level of regulation of current manual CCTV capabilities varies significantly between countries and between types of place. In many jurisdictions, only public authorities can deploy cameras in public places, although in others cameras placed on private property but which view public areas are permitted. Covert surveillance deployment in private or semi-private locations is also variable as to its legality and restrictions. As capabilities for automated processing of the footage from these systems become available, and then no doubt become cheaper, their use in workplaces, for example, is likely to rise. Existing data protection laws may offer some protections (such as the requirement for revelation of automated processing of personal data which effects job evaluation), but as shown by the UK's current position on the position of CCTV data as personal data (Adams, 2007) it would be better were this new area formally and clearly regulated, rather than relying on generic legislation.

The deployment of new PETs in advanced CCTV systems would seem to require government mandate rather than a reliance on deployer choice. An existing system option to overwrite individuals with “stick figures” offered by a Japanese security systems company was reported (Adams *et al.*, 2012) as discontinued because of lack of take-up by customers. Those paying for the deployment of CCTV systems are unlikely to pay extra for facilities which bring them no benefit, even indeed being seen as reducing the effectiveness of their investment in such systems, without government requirement that they do so in order to gain authority to deploy.

The deployment of these technologies will no doubt lead to their use in criminal investigations. In such cases issues of false positives in either identification or interpretation potentially carry even worse consequences for the falsely accused than such circumstances involving private actors’ systems. The use of such systems for tracking protesters is an obvious application, and likely to be used. Such technologies in use at violent protests in democracies might or might not raise serious questions of political freedoms. Their use in countries with obviously oppressive regimes would be more troubling. The reaction of various regimes in the Middle East to protests starting in Spring 2011 (and continuing in Syria particularly at the time of writing) have been alleged to include targetting the families of protesters for threats, arrest or punishments. Bans on the export of these technologies to oppressive regimes should certainly be considered. Requirements of release of all relevant video footage and the making available of equivalent automated assessment to the defence in criminal trials as are used by the prosecution should also be under consideration.

These technologies are under swift development, often with public funds. A debate about their use, and potential misuse, abuse and possible unintended consequences is needed.

References

Adams, A. A. 2007. Regulating CCTV. *Pages 3–14 of: Ethicomp 2007: Proceedings of the Ninth International Conference*. Tokyo: Meiji University.

Adams, A. A., Murata, K., & Orito, Y. 2010. The Development of Japanese Data Protection. *Policy and Internet*, **2**(2), 95–126. Article 5.

Adams, A. A., Murata, K., & Orito, Y. 2012. *Tender Electronic Eyes: Case Studies on CCTV Deployment in Japan*. In Preparation.

Annesley, J., Colombo, A., Orwell, J. and Velastin, S. A. 2007. A Profile of MPEG-7 for Visual Surveillance. *In Proceedings Fourth IEEE Advanced Video and Signal-Based Surveillance (AVSS), 482-487.*

BBC. 2003 (February). *Pensioner Freed After FBI Bungle*. news.bbc.co.uk/2/hi/uk_news/england/2799791.stm

BBC. 2006 (January). *Peeping Tom CCTV Workers Jailed*. news.bbc.co.uk/2/hi/uk_news/england/merseyside/4609746.stm

BBC. 2012 (October 3). *High-Def CCTV Cameras Risk Backlash, Warns UK Watchdog*. www.bbc.co.uk/news/technology-19812385

Bazzani, L., Cristani, M., Perina, M. & Murino, V. 2012. Multiple-Shot Person Re-Identification by Chromatic and Epitomic Analyses. *In Pattern Recognition Letters*, **33**(7), 898-903.

Ben Shitrit, H., Berclaz, J., Fleuret, F. & Fua, P. 2011. Tracking Multiple People under Global Appearance Constraints, *Computer Vision (ICCV), 2011 IEEE International Conference on*, pp.137-144.

Berger, A, M. 2000. Privacy Mode for Acquisition Cameras and Camcorders. U.S. Patent 6,067,399, Sony Corporation, 23 May 2000.

Brin, D. 1998. *The Transparent Society*. Jackson, TN: Perseus.

Cezar, J., Jacques, S., & Musse, S. R. 2010. Crowd Analysis Using Computer Vision Techniques. *IEEE Signal Processing Magazine*, **27**(September), 66–77.

Coetzer, B., Merwe, J. van der, & Josephs, B. 2011. *Information Management and Video Analytics: the Future of Intelligent Video Surveillance*. InTech. http://tinyurl.com/78flxuy

Coudert, F., Gemo, M., Beslay, L., & Andritsos, F. (2011). Pervasive Monitoring: Appreciating Citizen's Surveillance as Digital Evidence in Legal Proceedings. In Imaging for Crime Detection and Prevention 2011 (ICDP 2011), 4th International Conference on (pp. 1-6).

Dee, H., & Hogg, D. 2004. Detecting Inexplicable Behaviour. *Pages 477–486 of: Proceedings of the British Machine Vision Conference, The British Machine Vision Association*.

Dee, H., & Hogg, D. 2006. Navigational Strategies and Surveillance. *Pages 73–81 of: Proceedings of the IEEE International Workshop on Visual Surveillance (ECCV-VS)*. IEEE Computer Society Press.

Dee, H. M. & Velastin, S. A. 2008. How Close Are We To Solving the Problem of Automated Visual Surveillance? A Review of Real-World Surveillance, Scientific Progress and Evaluative Mechanisms. *Machine Vision and Applications*, **19**(5-6), 329-343.

Dick, A. R., & Brooks, M. J. 2003. Issues in Automated Visual Surveillance. *In:* Sun *et al.* (2003).

Dollár, P., Wojek, C., Schiele, B. & Perona, P. 2012. Pedestrian Detection: An Evaluation of the State of the Art, *In Pattern Analysis and Machine Intelligence (PAMI)*, **34**(4), 743-761.

Donald, C. 2005. How Many Monitors Should a CCTV Operator View? In *CCTV Image*. STL Publishing, London, England, pp. 355.

Dufaux, F. & Ebrahimi, T. 2008. Scrambling for Privacy Protection in Video Surveillance Systems. *In IEEE Transactions on Circuits and Systems for Video Technology*, **18**(8), 1168-1174.

Evans, M., Boyle J. N. & Ferryman, J. 2012, Vehicle Classification using Evolutionary Forests. *Proceedings International Conference on Pattern Recognition Applications and Methods*, pp 387-393.

Florence, J. 2006. Making the No Fly List Fly: A Due Process Model for Terrorist Watchlists. *Yale Law Journal*, **115**(8), 2148–2181. www.yalelawjournal.org/pdf/115-8/Florence.pdf.

Gong, S., Loy, C. C., & Xiang, T. 2011. *Security and Surveillance*. Springer.

Goold, B. J. 2002. Privacy Rights and Public Spaces: CCTV and the Problem of the "Unobservable Observer". *Criminal Justice Ethics 21(1) 21-27.*

Helbing, D., & Molnar, P. 1995. Social Force Model for Pedestrian Dynamics. *Physical Review E*, **51**.

Hempel, L., & Töpfer, E. 2004. *CCTV in Europe (Final Report)*. Tech. rept. Urban Eye Project: HPSE-CT2001-00094.

Hu, W., Tan, T., Wang, L., & Maybank, S. 2004. A Survey on Visual Surveillance of Object Motions and Behaviour. *IEEE Transactions on Systems, Man, and Cybernetics*, **34**, 334–252.

Independent Police Complaints Commission (UK). 2007 (8 November). *Stockwell One: Investigation into the shooting of Jean Charles de Menezes at Stockwell underground station on 22 July 2005*. www.ipcc.gov.uk/Documents/stockwell_one.pdf.

Introna, L. D. (2000). Workplace Surveillance, Privacy and Distributive Justice. ACM SIGCAS Computers and Society, 30(4), 33-39.

Introna, L. D., & Wood, D. 2004. Picturing Algorithmic Surveillance: The Politics of Facial Recognition Systems. *Surveillance and Society*, **2**(2/3), 177–198.

Jia, K. & Gong, S. 2008. Generalized Face Super-Resolution. *In IEEE Transactions on Image Processing*, **17**(6), 873-886.

Ko, T. 2011. *Video Surveillance*. InTech. Chap. A Survey on Behaviour Analysis in Video Surveillance Applications. http://tinyurl.com/8hlveuk

Lee, S-W., Park, J., & Lee, S-W. 2006. Low Resolution Face Recognition Based on Support Vector Data Description. *Pattern Recognition*, **39**, 1809–1812.

Lewis, P. 2009. G20 protests death Ian Tomlinson death: Guardian video reveals police attack on man who died at G20 protest. *The Guardian*, 7th April. www.guardian.co.uk/uk/2009/apr/07/ian-tomlinson-g20-death-video.

Liu, C., Shum, H-Y., & Freeman, W. T. 2007. Face Hallucination: Theory and Practice. *International Journal of Computer Vision*, **75**(1), 115–134.

Lyall, S. 2009. Critics Assail British Police for Harsh Tactics During the G-20 Summit Meeting. *New York Times*, 30th May. www.nytimes.com/2009/05/31/world/europe/31police.html.

MarketResearch.com. 2011 (January). *Global Video Surveillance Market, Applications and Management Services Forecasts (2010-2015).* www.marketresearch.com/MarketsandMarkets-v3719/Global-Video-Surveillance-Applications-Management-6083117/.

Mehran, R., Moore, E. R., & M, Shah. 2010. A Streakline Representation of Flow in Crowded Scenes. *In: European Conference on Computer Vision (ECCV)*.

Moussaid, M., Perozo, N., Garnier, S., Helbing, D., & Theraulaz, G. 2010. The Walking Behaviour of Pedestrian Social Groups and Its Impact on Crowd Dynamics. *PloS ONE*, **5**(4).

Norris, C., & Armstrong, G. 1999. *The Maximum Surveillance Society: The Rise of CCTV*. Oxford: Berg.

Orrite, C., Martinez, F., Herrero, E., Ragheb, H. & and Velastin, S. 2008. Independent Viewpoint Silhouette-based Human Action Modelling and Recognition, *Proc. 1st*

International Workshop on Machine Learning for Vision-based Motion Analysis in conjunction with the European Conference on Computer Vision (ECCV).

Park, J-S., & Lee, S-W. 2008. An Example-Based Face Hallucination Method for Single-Frame, Low-Resolution Facial Images. *IEEE Transactions on Image Processing*, **17**(10), 1806–1816.

Patino, L., Bremond, F., Evans, M., Shahrokni, A. & Ferryman, J. 2010. Video Activity and Reporting with Incremental Unsupervised Learning. *In Proceedings Seventh IEEE Advanced Video and Signal-Based Surveillance (AVSS),* 511-518.

Ragheb, H., Velastin, S., Remagnino, P. & Ellis, T. 2008. ViHASi: Virtual Human Action Silhouette Data for the Evaluation of Silhouette-Based Action Recognition Methods, *Proc. ACM International Workshop on Vision Networks for Behaviour Analysis in conjunction with the ACM Multimedia (MM)*.

PETS2006. 2006. *IEEE 2006 International Workshop on Performance Evaluation of Tracking and Surveillance*. www.pets2006.net.

PETS2007. 2007. *IEEE 2007 International Workshop on Performance Evaluation of Tracking and Surveillance*. www.pets2007.net.

PETS2009. 2009 *IEEE 2009 International Workshop on Performance Evaluation of Tracking and Surveillance*. www.pets2009.net.

Pham, Q-C., Gond, L., Begard, J., Allezard, N. & Sayd, P. Real Time Posture Analysis in a Crowd using Thermal Imaging. *In Proceedings Computer Vision and Pattern Recognition (CVPR'07*), pp 1-8.

Reeves, B., & Nass, C. 2002. *The Media Equation*. 2nd edn. Stanford: CSLI.

Rowlands, M. 2011 (21 January). *Statewatch Analysis — Six Months On: An Update on the UK Coalition Government's Commitment to Civil Liberties*. www.statewatch.org/analyses/no-118-uk-civil-liberties-six-months-on.pdf.

Senior, A., Pankanti, A., Brown, L., Ying-Li, T., Ekin, A., Connell, J., Chiao, F.-S., & Lu, M. 2005. Enabling Privacy through Computer Vision. *IEEE Security and Privacy*, **3**, 50–57.

Sochman, J. & Hogg, D. C. 2011. Who Knows Who – Inverting the Social Force Model for Finding Groups. *In Proceedings IEEE International Workshop on Socially Intelligent Surveillance and Monitoring (SISM)*, pp 830-837.

Spindler, T., Wartmann, C., Hovestadt, L., Roth, D., Van Gool, L., & Steffen, A. 2008. Privacy in Video Surveilled Spaces. *Computer Security*, **16**(2), 199–222.

Sun, C., Talbot, H., Ourselin, S., & Adriaansen, T. (eds). 2003. *Proceedings of the Seventh International Conference on Digital Image Computing: Techniques and Applications, DICTA 2003, 10-12 December 2003, Macquarie University, Sydney, Australia*. CSIRO Publishing.

The Economist. 2012 (28 April). Video Surveillance: I Spy with My Big Eye. www.economist.com/node/21553408

Tickner, A.H. & Poulton, E.C. 1973. Monitoring up to 16 Synthetic Television Pictures Showing a Great Deal of Movement. *Ergonomics*, *16*, 381-401.

Tweed, D. & Ferryman, J. 2008. Enhancing Change Detection in Low-Quality Surveillance Video with Markov Random Fields", *Proc. 1st ACM International Workshop on Vision Networks for Behaviour Analysis (VNBA'08) (in conjunction with the ACM Multimedia Conference)*.

Valera, M., Velastin, S., Ellis A-L., & Ferryman, J. 2011. Communication Mechanisms and Middleware for Distributed Video Surveillance, *IEEE Transactions on Circuits and Systems for Video Technology,* **21**(12), 1795-1809.

Welsh, B. P. and Farrington, D. P. 2009. Public Area CCTV and Crime Prevention: An Update Systematic Review and Meta-Analysis. *Justice Quarterly 26(4), 716-745.*

Yilmaz, A., Javed, O., & Shah, M. 2006. Object Tracking: A Survey. *ACM Computing Surveys*, **38**(4), Article 13.

Zaidenberg, S., Boulay, B., & Bremond, F. 2012. A Generic Framework for Video Understanding Applied to Group Behaviour Recognition, *Proceedings of the IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance (AVSS 2012), 126-142.*

Zhan, B., Monekosso, D. N., Remagnino, P., Velastin, S. A., & Xu, L. Q. 2008. Crowd Analysis: A Survey. *Mach. Vis. Appl.*, **19**(5-6), 345–357.